BAYESIAN DESIGN AND ANALYSIS OF CLUSTER RANDOMIZED TRIALS

Shan Xiao

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Biostatistics,

Indiana University

October, 2017

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Wanzhu Tu, Ph.D., Co-Chair

Doctoral Committee

Ziyue Liu, Ph.D., Co-Chair

Spencer George Lourens, Ph.D.

August 7, 2017

Huanmei Wu, Ph.D.

DEDICATION

To My Beloved Family.

ACKNOWLEDGMENTS

I would like to thank Drs. Wanzhu Tu and Ziyue Liu, my co-advisors, for their guidance and support throughout my dissertation training. I am very fortunate to have Dr. Tu as my mentor for the past three years. His commitment to research and pursuit of quality work have set an excellent example for me. He has always been willing to help me without reservation and to give me professional advice with his exceptional insight. From him, I learned not only how to do research in Biostatistics, but also how to be a scholar.

My sincere appreciation goes to Drs. Spencer George Lourens and Huanmei Wu. It has been a great learning experience to work with Dr. Lourens, especially on developing the R package. He has guided me through various challenges in computing related matters. Finally, I am grateful to Dr. Gregory D. Zimet for allowing me to test my method using his data.

I'd also like to thank all faculty members in the Biostatistics Program for creating this stimulating academic environment. I appreciate the support from staff members in the department. Finally, I would like to thank my fellow students for their continuous support and encouragement, which have made my graduate study a wonderful experience.

Shan Xiao

BAYESIAN DESIGN AND ANALYSIS OF CLUSTER RANDOMIZED TRIALS

Cluster randomization is frequently used in clinical trials for convenience of interventional implementation and for reducing the risk of contamination. The operational convenience of cluster randomized trials, however, is gained at the expense of reduced analytical power. Compared to individually randomized studies, cluster randomized trials often have a much-reduced power. In this dissertation, I consider ways of enhancing analytical power with historical trial data. Specifically, I introduce a hierarchical Bayesian model that is designed to incorporate available information from previous trials of the same or similar interventions. Operationally, the amount of information gained from the previous trials is determined by a Kullback-Leibler divergence measure that quantifies the similarity, or lack thereof, between the historical and current trial data. More weight is given to the historical data if they more closely resemble the current trial data. Along this line, I examine the Type I error rates and analytical power associated with the proposed method, in comparison with the existing methods without utilizing the ancillary historical information. Similarly, to design a cluster randomized trial, one could estimate the power by simulating trial data and comparing them with the historical data from the published studies. Data analytical and power simulation methods are developed for more general situations of cluster randomized trials, with multiple arms and multiple types of data following the exponential family of distributions. An R package is developed for practical use of the methods in data analysis and trial design.

Wanzhu Tu, Ph.D., Co-Chair

Ziyue Liu, Ph.D., Co-Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Background of the Research

Cluster randomization is widely used in clinical trials. The design allows for testing of interventional efficacy by comparing outcomes of individuals randomized by predefined subject clusters or groups (Campbell, 2000; Campbell et al., 2007). In practice, these clusters could be communities, families, or clinics. With this design, clusters participating in a trial are randomly assigned to different treatment arms, and all subjects within each cluster receive the same treatment. Trial outcomes are typically assessed at the subject level, as in individual randomized trials.

Trialists usually choose the cluster randomized design for practical reasons (Donner, 1998), such as convenience of implementation, or minimization of contamination, etc. The design is particularly popular in situations where the intervention is delivered at the cluster level. For example, in trials of electronic care reminders, it is usually more practical to deliver the same reminder message to all patients in a clinic, as opposed to sending different messages to different patients, or withholding messages from randomly selected patients. In such situations, individual randomization creates not only logistic and implementational difficulties, but also a potential to contaminate the treatment effect through interactions among patients within the same clinic.

But the practical appeal of cluster randomization is retained at the expense of a much reduced analytical power. In a typical two-arm cluster randomized trials with k clusters per treatment arm and m subjects per cluster (i.e., a total of $n_C = km$ subjects per arm), for example, one would need VIF= $[1 + (m - 1)\rho]$ times of the sample size of an individual randomized trial to detect an effect of the same magnitude at the same power and level of significance (Hemming et al., 2011). Here ρ indicates the intra-cluster correlation (ICC) coefficient, and VIF stands for Variance Inflation Factor. As a result, in situations where there are relatively few large clusters with a strong ICC, the inflation of sample size from individual randomized trials could be severe, leading to a great reduction in analytical power.

How to improve analytical power in cluster randomized trials, therefore, becomes a question of great practical importance. In this proposal, we present a class of Bayesian methods for the design and analysis of cluster randomized trials. The proposed methods enhance the analytical power of cluster randomized trials by borrowing strength from historical data.

1.2 Enhancing Power with Historical Data

The idea of borrowing information from historical data is not new. Using information from previous trials of similar interventions to boost the power of the current trial is intuitively an appealing idea (Viele et al., 2014). After all, efficacies of many interventions are established through repeated testing in multiple trials, in different patient populations and clinical settings, and through different delivery methods.

Philosophically, such an approach is no different from meta analysis, which seeks to quantify an unknown treatment effect by combining all existing trials on the

same or similar interventions (Donner and Klar, 2002; Donner et al., 2001). While people continue to debate the perceived fault of ignoring the inherent heterogeneity among the historical trials, few challenge the notion that data from multiple studies could be combined to achieve an improved treatment effect estimate (Darlington and Donner, 2007; Donner et al., 2003; Laopaiboon, 2003; Shuster et al., 2007). In fact, systematic review and meta analysis are generally regarded as at the top of the hierarchy of scientific evidence (Guyatt et al., 1995).

Alternatively, one could approach the problem from the perspective of hierarchical Bayes, by formulating the prior distributions based on historical data (Clark and Bachmann, 2010; Spiegelhalter, 2001; Turner et al., 2001). Specifically, a frequently used method is to determine the prior parameters from historical data (Goodman and Sladky, 2005; Hampson et al., 2014; Hobbs and Carlin, 2007; Schoenfeld et al., 2009; Turner et al., 2005). To prevent an overwhelming influence of the historical data, several researchers developed the idea of "discounting" the historical information. Magnitude of the discount is quantified by a power parameter associated with the prior density, and thus leading to the concept of power prior (Duan, 2005; Ibrahim et al., 2015; Zhang, 2010).

1.3 An Overview of the Dissertation Research

An essential question concerning the use of prior trial data is how to determine the relevance of historical information. For simplicity of the discussion, we assume that those previous studies are on the same intervention. But having the same intervention does not automatically give the historical data the same weight as the current trial data, because the designer of the current trial has no control of the target populations,

the settings, and the conduct of the previous trials. In other words, findings from the previous studies are not necessarily generalizable to the current population and clinical setting.

In the absence of a generally accepted mechanism for determination of the similarity or dissimilarity between the current and historical trials, a reasonable approach is to adopt a data-driven method that directly evaluates the resemblance of the distributions of the data sources. A greater resemblance indicates a stronger affinity between the historical and current trials. A lower resemblance, on the other hands, conveys a sense of dissimilarity.

In this dissertation, we propose to use the Kullback-Leibler (KL) divergence measure to quantify the distance between the current and historical data, and then use this distance measure to determine the amount of discounting of historical data. A greater KL distance indicates a bigger discrepancy, and thus giving us less incentive to place great weight on the historical information. To implement, we propose to use the KL distance as the discounting parameter in a likelihood framework.

Our research has three separate but related pieces: (1) Bayesian Analysis of data from cluster randomized trials; (2) Design of cluster randomized trials with historical data; and (3) Development of computational tools for the data analysis and trial design.

The methods that we present in this dissertation are Bayesian in the sense that they use Bayesian techniques for incorporation of the historical data. The basic clinical trial concepts, such as type I error rate and power, however, are still within the frequentist trial framework. As we shall describe in greater details, such a merge of the Bayesian and frequentist views offers an enhanced modeling flexibility, especially when

it comes to the incorporation of extraneous historical trial information, but at the same time, it retains the more familiar frequentist trial concepts and interpretations.

Chapter 2

Analysis of Two-arm Cluster Randomized Trials with Binary Outcome

2.1 Bayesian Power Priors

For a given trial, all information provided by previous studies are ancillary in nature. One way to incorporate such information is through the use of power priors (Ibrahim et al., 2015). Bayesian power prior has been studied by multiple authors (Chen et al., 2006, 2000, 1998; De Santis, 2006, 2007; Duan et al., 2006; Ibrahim and Chen, 2000; Matteucci and Veldkamp, 2015; Rietbergen et al., 2011; Shao, 2012). To the best of my knowledge, however, it has not been used in cluster randomized trials. In this work, we briefly review and describe how power priors are used to bring prior trial data into the analysis. We start the discussion by considering the simplest situation of one historical study.

We write the current data as $\boldsymbol{D}$ and historical data as $\boldsymbol{D}_0$. We additionally assume that the likelihood functions based on the current and historical data are $L(\boldsymbol{\theta}|\boldsymbol{D})$ and $L(\boldsymbol{\theta}|\boldsymbol{D}_0)$, respectively, where $\boldsymbol{\theta}$ is the parameter of interest. With a being a discounting parameter, the power prior of $\boldsymbol{\theta}$ is constructed as follows,

$$\pi(\boldsymbol{\theta}|\boldsymbol{D_0}, a) \propto L(\boldsymbol{\theta}|\boldsymbol{D_0})^a \pi_0(\boldsymbol{\theta}), \tag{2.1}$$

where $\pi_0(\boldsymbol{\theta})$ is the original prior before the historical data $\boldsymbol{D}_0$ are made available. In practice, $\pi_0(\boldsymbol{\theta})$ is typically chosen to be a non-informative probability density function with heavier tails.

The discounting parameter a takes a value between 0 and 1. When $a = 0$, the prior (2.1) degenerates to $\pi_0(\boldsymbol{\theta})$, which means all information in the historical data is discounted and no ancillary information is incorporated in the prior specification; when $a = 1$, (2.1) represents a distribution generated by updating $\pi_0(\boldsymbol{\theta})$ with the full historical information through the Bayes theorem, implying no historical data will be discounted. In general, the discounting fraction a controls the levels of influence of historical data in the prior distribution formulation. Technically, parameter a controls the shape of the power prior function (2.1). As a becomes smaller, the tails of the power prior become heavier.

The basic formulation of the power prior is easily extendable to situations of multiple historical data sources. Suppose there are J datasets from historical studies, $\boldsymbol{D}_1, \ldots, \boldsymbol{D}_J$, each of which corresponds to a unique discounting factor $a_1, \ldots, a_J$, the power prior can therefore be expressed as

$$\pi(\boldsymbol{\theta}|\boldsymbol{D_1}, \ldots, \boldsymbol{D_J}, a_1, \ldots, a_J) \propto \pi_0(\boldsymbol{\theta}) \prod_{j=1}^{J} L(\boldsymbol{\theta}|\boldsymbol{D}_j)^{a_j}. \tag{2.2}$$

For simplicity of the narration, we shall limit the discussion to the situation of one historical data, knowing that the method is easily extendable to situations where multiple historical data sources are available.

The power prior in Equation (2.1) can be updated with the current data likelihood $L(\boldsymbol{\theta}|\boldsymbol{D})$ via the Bayes theorem, resulting in a posterior of $\boldsymbol{\theta}$ in the following

form,

$$\pi(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{D_0}, a) \propto L(\boldsymbol{\theta}|\boldsymbol{D})\pi(\boldsymbol{\theta}|\boldsymbol{D_0}, a) \propto L(\boldsymbol{\theta}|\boldsymbol{D})L(\boldsymbol{\theta}|\boldsymbol{D_0})^a\pi_0(\boldsymbol{\theta}). \tag{2.3}$$

One could use posterior (2.3) to estimate $\boldsymbol{\theta}$, and perform inference.

It should be noted that Equation (2.1) assumes that the parameter of interest is identical across all trials (Ibrahim and Chen, 2000). Such an assumption can be relaxed by allowing the main parameter of interest to differ across trials, as long as it follows a distribution that depends only on a common set of parameters from the historical trials (Hobbs et al., 2011, 2012). An alternative way of accommodating varying parameters is to assume the parameters are exchangeable in the current and historical trials (Carlin, 1992). Such an assumption is frequently used in Bayesian meta analysis, but rarely used in Bayesian power prior analysis.

2.2 Determination of Discounting Parameters

An essential step to incorporate the previous trial data is determining an appropriate value for the discounting parameter a. One of the common methods is to subjectively determine the value of a; the practice is usually based on expert opinions (De Santis, 2007; Rietbergen et al., 2011; Shao, 2012). To implement, analysts often try different values between 0 and 1, and report results to content experts. The final decisions are based on certain pre-specified criteria, such as the marginal likelihood criterion (MLC) and deviance information criterion (DIC) (Ibrahim et al., 2015). Alternatively, one could treat a as a random variable, and specify a hyper-prior to describe its behavior, which is usually described by a non-informative distribution or a distribution driven by a commensurability parameter (Brian, 2010; Hobbs et al., 2011;

Ibrahim et al., 2015). This latter approach is conceptually more attractive as the discounting fraction is essentially data-driven. A downside of the approach is that it often requires extensive computation.

Suppose that we have a two-arm cluster randomized trial with a binary outcome, and the historical data come from a two-arm randomized trial on a similar outcome. The historical trial is not necessarily cluster randomized, and its data may come in a summarized fashion.

For the current trial, we assume the numbers of clusters in the control and treatment arms to be I_0 and I_1, respectively. So the total number of clusters is $I = I_0 + I_1$. Let J_i be the number of subjects in the ith cluster and Y_{ij} be the binary outcome for the jth subject in the ith cluster. Y_{ij} takes the value 1 if the event of interest is observed and 0 otherwise. Treatment assignment of the ith cluster is indicated by X_{1i}. The current trial data are therefore written as $\boldsymbol{D} = \{(Y_{ij}, X_{1i}), i = 1, 2, \ldots, I, j = 1, 2, \ldots, J_i\}$.

For the historical study, we let l be the treatment group indicator, with $l = 0$ indicating the control arm and $l = 1$ the treatment arm. Assuming that the data come in a summarized form, and that N_l indicates the total number of subjects in arm l, we let Z_l be the number of events of interest observed in the lth arm and let X_{2l} be the treatment assignment. Then the historical data can be summarized as $\boldsymbol{D}_0 = \{(Z_l, N_l, X_{2l}) : l = 0, 1\}$.

To determine the discounting fraction a, we propose the following models for $\boldsymbol{D}$ and $\boldsymbol{D}_0$, which can be fitted separately (Congdon, 2007). Specifically, for the

current data $\boldsymbol{D}$, we assume $Y_{ij} \sim \text{Bernoulli}(p_i)$, and

$$\text{logit}(p_i) = \beta_0 + \beta_1 X_{1i} + b_i, i = 1, 2, \ldots, I, j = 1, 2, \ldots, J_i, \tag{2.4}$$

where $b_i \sim \text{N}(0, \tau_b)$, $\beta_0 \sim \text{N}(\mu_{\beta_0}, \tau_{\beta_0})$, $\beta_1 \sim \text{N}(\mu_{\beta_1}, \tau_{\beta_1})$, $\tau_b \sim \text{Gamma}(\kappa_b, \nu_b)$, and β_1 represents the treatment effects under $\boldsymbol{D}$. For the simplicity of narration, we denote $\boldsymbol{b} = (b_1, b_2, \ldots, b_I)^T$. Similarly, for the historical data $\boldsymbol{D}_0$, we assume $Z_l \sim \text{Binomial}(N_l, q_l)$, and

$$\text{logit}(q_l) = \beta_0' + \beta_1' X_{1l}', l = 0, 1, \tag{2.5}$$

where $\beta_0' \sim \text{N}(\mu_{\beta_0'}, \tau_{\beta_0'})$, $\beta_1' \sim \text{N}(\mu_{\beta_1'}, \tau_{\beta_1'})$ and β_1' stands for the treatment effect under $\boldsymbol{D}_0$.

All priors for parameters in models (2.4) and (2.5) are assumed to be non-informative. In separate data analyses, the posterior of β_1 and β_1', denoted by $f(\beta_1|\boldsymbol{D})$ and $g(\beta_1'|\boldsymbol{D}_0)$, are evaluated.

Since the two posteriors are derived under non-informative priors, they are not influenced by external data other than the current and historical trial information. From this, we ascertain the symmetric and asymmetric KL divergence measures, and use them to quantify the similarity of the two data sources.

$$D_{\text{KL}}^{\text{sym}}(f||g) = \text{E}_f\{\log(\frac{f}{g})\} + \text{E}_g\{\log(\frac{g}{f})\}, \tag{2.6}$$

$$D_{\text{KL}}^{\text{asym}}(f||g) = \text{E}_f\{\log(\frac{f}{g})\}, \tag{2.7}$$

where f and g stand for $f(\beta_1|\boldsymbol{D})$ and $g(\beta_1'|\boldsymbol{D}_0)$, and $\mathrm{E}_f\{\cdot\}$ and $\mathrm{E}_g\{\cdot\}$ are the expectations taken with respect to f and g, respectively.

Under such a setup, we propose the following discounting fractions

$$a^{\text{sym}} = e^{-D_{KL}^{\text{sym}}(f||g)} = e^{-D_{KL}^{\text{sym}}(g||f)}. \tag{2.8}$$

$$a^{\text{asym}} = e^{-D_{KL}^{\text{asym}}(f||g)}, \tag{2.9}$$

The two KL divergence measures are not easy to compute by definition, we therefore use the k-Nearest Neighbor (k-NN) algorithm for calculation.

2.3 Estimation of Treatment Effect

We assume $\boldsymbol{\beta}_1 = (\beta_1, \beta_1')$ follows a bivariate normal distribution with mean vector $\boldsymbol{m_1}$ and precision matrix $\boldsymbol{\Lambda_1}$. Assuming β_1 and β_1' are exchangeable, we could write $\boldsymbol{m_1} = \mu_1(1,1)'$, and $\boldsymbol{\Lambda_1} = \tau_1 \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}^{-1}$. To complete the model specification, we further assume μ_1, τ_1 and ρ_1 follow the hyper-prior distributions as below,

$$h_1(\mu_1|a_1, R_1) = \sqrt{\frac{R_1}{2\pi}} e^{-\frac{R_1}{2}(\mu_1 - a_1)^2},$$

$$h_1(\tau_1|\kappa_1, \nu_1) = \frac{\nu_1^{\kappa_1}}{\Gamma(\kappa_1)} \tau_1^{\kappa_1 - 1} e^{-\nu_1 \tau_1}, \tag{2.10}$$

$$h_1(\rho_1|c_1, d_1) = \frac{\rho_1^{c_1 - 1}(1-\rho_1)^{d_1 - 1}}{B(c_1, d_1)},$$

where $\Gamma(\cdot)$ and $B(\cdot,\cdot)$ are respectively Gamma and Beta functions. The power prior of β_1 can therefore be written as

$$h(\beta_1|\boldsymbol{D}_0,a) \propto \iiiint \{\int (L(\beta_0',\beta_1'|\boldsymbol{D}_0))^a g(\beta_0')\,d\beta_0'\} h(\beta_1,\beta_1'|\mu_1,\tau_1,\rho_1) \\ h(\mu_1|a_1,R_1)h(\tau_1|\kappa_1,\nu_1)h(\rho_1|c_1,d_1)\,d\beta_1'\,d\mu_1\,d\tau_1\,d\rho_1, \tag{2.11}$$

where $L(\beta_0',\beta_1'|\boldsymbol{D}_0)$ and $g(\beta_0')$ are the likelihood and prior density functions of β_0' and β_1' in Model (2.5). From (2.11), we derive the posterior function of β_1 as

$$h(\beta_1|\boldsymbol{D},\boldsymbol{D}_0,a) \propto \{\int \cdots \int L(\beta_0,\beta_1|\boldsymbol{D},\boldsymbol{b}) f(\boldsymbol{b}|\tau_b) f(\beta_0) f(\tau_b)\,d\boldsymbol{b}\,d\beta_0\,d\tau_b\} h(\beta_1|\boldsymbol{D}_0,a), \tag{2.12}$$

where $L(\beta_0,\beta_1|\boldsymbol{D},\boldsymbol{b})$, $f(\beta_0)$,$f(\tau_b)$ and $f(\boldsymbol{b}|\tau_b)$ are the (conditional) likelihood and (prior) density functions of β_0, β_1, τ_b and $\boldsymbol{b}$ in Model (2.4).

Since the three posteriors $f(\beta_1|\boldsymbol{D})$, $g(\beta_1'|\boldsymbol{D}_0)$, $h(\beta_1|\boldsymbol{D},\boldsymbol{D}_0,a)$ do not have closed forms, we use a Metropolis Hastings (MH) algorithm within Gibbs sampling to draw samples from the posteriors. Treatment effect estimates are then obtained from the appropriate summary statistics of the posterior samples.

2.4 Algorithms

2.4.1 The K-Nearest Neighbor Algorithm

To estimate the symmetric and asymmetric KL divergence (2.7) and (2.6) with the posterior samples of β_1 from $f(\beta_1|\boldsymbol{D})$ and β_1' from $g(\beta_1'|\boldsymbol{D}_0)$, people usually obtain the histograms of equally sized bins with the posterior samples and then use these histograms to estimate the posterior densities f and g and substitute the density

estimates $\hat{f}$ and $\hat{g}$ into equations (2.7) and (2.6). But this method is extensive in computation, and the estimation accuracy deteriorates quickly as the dimension of the densities f and g increases. To solve, an estimator of the KL divergence with the nearest neighbor method has been proposed (Wang et al., 2009). This estimator demonstrates asymptotically unbiased and mean-square consistent properties when the posterior sample size is large enough. In this dissertation, the k-Nearest Neighbor (k-NN) algorithm is applied to the posterior samples to yield this estimator, where k is any constant since different k do not affect the two aforementioned properties but the convergence rate of the k-NN algorithm. The R package 'FNN' (Beygelzimer et al., 2013) is used to complete the computations. For the purpose of illustration, the application of the k-NN algorithm in this dissertation is briefly introduced.

Let $\{\beta_1^{(1)}, \beta_1^{(2)}, \ldots, \beta_1^{(n)}\}$ and $\{\beta_1'^{(1)}, \beta_1'^{(2)}, \ldots, \beta_1'^{(m)}\}$ be the posterior samples of β_1 from $f(\beta_1|\boldsymbol{D})$ and β_1' from $g(\beta_1'|\boldsymbol{D}_0)$, respectively. Let $\rho_k(i)$ be the Euclidean distance between $\beta_1^{(i)}$ and its k-NN in $\{\beta_1^{(j)}\}_{j\neq i}$. Specifically, the k-NN of x in $\{z_1, z_2, \ldots, z_n\}$ is $z_{i(k)}$ where $i(1), i(2), \ldots, i(n)$ is such that

$$||x - z_{i(1)}|| \leq ||x - z_{i(2)}|| \leq \cdots \leq ||x - z_{i(n)}||.$$

The distance from $\beta_1^{(i)}$ to its k-NN in $\{\beta_1'^{(1)}, \beta_1'^{(2)}, \ldots, \beta_1'^{(m)}\}$ is denoted by $\nu_k(i)$, then the asymmetric KL divergence $D_{\mathrm{KL}}^{\mathrm{asym}}(f||g)$ can be calculated as

$$\hat{D}_{\mathrm{KL}}^{\mathrm{asym}}(f||g) = \frac{1}{n}\sum_{i=1}^{n} \log(\frac{\nu_k(i)}{\rho_k(i)}) + \log(\frac{m}{n-1}), \tag{2.13}$$

Similarly, let $\kappa_k(j)$ be the Euclidean distance between $\beta_1'^{(j)}$ and its k-NN in $\{\beta_1'^{(i)}\}_{i\neq j}$, and the Euclidean distance from $\beta_1'^{(j)}$ to its k-NN in $\{\beta_1^{(1)}, \beta_1^{(2)}, \ldots, \beta_1^{(n)}\}$ is denoted

by $\lambda_k(j)$, then the symmetric KL divergence $D_{\mathrm{KL}}^{\mathrm{sym}}(f||g)$ can be calculated as

$$\hat{D}_{\mathrm{KL}}^{\mathrm{sym}}(f||g) = \frac{1}{n}\sum_{i=1}^{n}\log(\frac{\nu_k(i)}{\rho_k(i)}) + \log(\frac{m}{n-1}) + \frac{1}{m}\sum_{j=1}^{m}\log(\frac{\lambda_k(j)}{\kappa_k(j)}) + \log(\frac{n}{m-1}). \quad (2.14)$$

Consequently, the symmetric and asymmetric discounting fraction could be estimated as

$$\hat{a}^{\mathrm{sym}} = e^{-\hat{D}_{\mathrm{KL}}^{\mathrm{sym}}(f||g)} = e^{-\hat{D}_{\mathrm{KL}}^{\mathrm{sym}}(g||f)}.$$

$$\hat{a}^{\mathrm{asym}} = e^{-\hat{D}_{\mathrm{KL}}^{\mathrm{asym}}(f||g)},$$

2.4.2 The Metropolis Hastings within Gibbs Algorithm

Markov Chain Monte Carlo (MCMC) method has been proposed to draw posterior samples from a posterior distribution (Robert and Casella, 2009) when the posterior distribution does not have a closed form. It is built upon two basic algorithms, i.e. Gibbs algorithm and MH algorithm. These two algorithms have been extended and modified in different ways for solving different problems (Chib and Carlin, 1999; Gamerman, 1997). Many user-friendly statistical softwares have been well developed to implement MCMC procedure, such as OpenBUGS (Lunn et al., 2009). The development of the softwares makes MCMC easy to implement and helps it gain popularity in Bayesian computation. To draw posterior samples from $f(\beta_1|\boldsymbol{D})$, $g(\beta_1'|\boldsymbol{D}_0)$ and $h(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0, a)$, we propose a MH within Gibbs algorithm, and we develop both R and OpenBUGS programs to run this algorithm.

The Gibbs Component of the Hybrid Algorithm

We fit the current data $\boldsymbol{D}$ with Model (2.4) to draw posterior samples of β_1 from the posterior $f(\beta_1|\boldsymbol{D})$. The joint posterior of β_1, β_0, $\boldsymbol{b}$ and τ_b is

$$
\begin{aligned}
&f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D}) = \frac{f(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\boldsymbol{D})}{f(\boldsymbol{D})}\\
&= \frac{f(\boldsymbol{Y}|\beta_1,\beta_0,\boldsymbol{b})f(\beta_1)f(\beta_0)f(\boldsymbol{b}|\tau_b)f(\tau_b)}{\int\cdots\int f(\boldsymbol{Y}|\beta_1,\beta_0,\boldsymbol{b})f(\beta_1)f(\beta_0)f(\boldsymbol{b}|\tau_b)f(\tau_b)\,d\beta_1\,d\beta_0\,d\boldsymbol{b}\,d\tau_b}\\
&\propto f(\boldsymbol{Y}|\beta_1,\beta_0,\boldsymbol{b})f(\beta_1)f(\beta_0)f(\boldsymbol{b}|\tau_b)f(\tau_b)\\
&= \left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\\
&\left\{\sqrt{\frac{\tau_{\beta_1}}{2\pi}}e^{-\frac{\tau_{\beta_1}}{2}(\beta_1-\mu_{\beta_1})^2}\right\}\left\{\sqrt{\frac{\tau_{\beta_0}}{2\pi}}e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\}\left\{\prod_{i=1}^{I}\sqrt{\frac{\tau_b}{2\pi}}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\frac{\nu_b^{\kappa_b}}{\Gamma(\kappa_b)}\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}\\
&\propto \left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\\
&\left\{e^{-\frac{\tau_\beta}{2}(\beta-\mu_\beta)^2}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\},
\end{aligned}
$$

where $\boldsymbol{Y} = (Y_{11},\ldots,Y_{1J_1},\ldots,Y_{I1},\ldots,Y_{IJ_I})$, $\boldsymbol{b} = (b_1,\ldots,b_I)$. So the full conditionals of β_1, β_0, $\boldsymbol{b}$ and τ_b are

$$
\begin{aligned}
&f(\beta_1|\boldsymbol{D},\beta_0,\boldsymbol{b},\tau_b) = f(\beta_1|\boldsymbol{D},\beta_0,\boldsymbol{b})\\
&= \frac{f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})}{f(\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})}\\
&= \frac{f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})}{\int f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})\,d\beta_1}\\
&= \frac{\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{e^{-\frac{\tau_{\beta_1}}{2}(\beta_1-\mu_{\beta_1})^2}\right\}}{\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{e^{-\frac{\tau_{\beta_1}}{2}(\beta_1-\mu_{\beta_1})^2}\right\}d\beta_1}\\
&\propto \left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{e^{-\frac{\tau_{\beta_1}}{2}(\beta_1-\mu_{\beta_1})^2}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&f(\beta_0|\boldsymbol{D},\beta_1,\boldsymbol{b},\tau_b) = f(\beta_0|\boldsymbol{D},\beta_1,\boldsymbol{b}) \\
&= \frac{f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})}{f(\beta_1,\boldsymbol{b},\tau_b|\boldsymbol{D})} \\
&= \frac{f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})}{\int f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})\,d\beta_0} \\
&= \frac{\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\}}{\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\}d\beta_0} \\
&\propto \left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&f(\boldsymbol{b}|\boldsymbol{D},\beta_1,\beta_0,\tau_b) = \frac{f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})}{f(\beta_1,\beta_0,\tau_b|\boldsymbol{D})} \\
&= \frac{f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})}{\int\cdots\int f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})\,d\boldsymbol{b}} \\
&= \frac{\left\{\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}}{\int\left\{\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}d\boldsymbol{b}} \\
&\propto \left\{\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&f(\tau_b|\boldsymbol{D},\beta_1,\beta_0,\boldsymbol{b}) = f(\tau_b|\boldsymbol{b}) \\
&= \frac{f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})}{f(\beta_1,\beta_0,\boldsymbol{b}|\boldsymbol{D})} \\
&= \frac{f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})}{\int f(\beta_1,\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})\,d\tau_b} \\
&= \frac{\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}}{\int\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}d\tau_b} \\
&\propto \left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\} \\
&\sim \text{Gamma}(\kappa_b+\frac{I}{2},\nu_b+\frac{\sum_i^I b_i^2}{2}).
\end{aligned}
$$

Secondly, we fit the historical data $\boldsymbol{D}_0$ with Model (2.5) to draw posterior samples of β_1' from the posterior $g(\beta_1'|\boldsymbol{D}_0)$. The joint posterior of β_1' and β_0' is

$$
\begin{aligned}
&g(\beta_1', \beta_0'|\boldsymbol{D}_0) = \frac{g(\beta_1', \beta_0', \boldsymbol{D}_0)}{g(\boldsymbol{D}_0)} \\
&= \frac{g(\boldsymbol{Z}|\beta_1', \beta_0')g(\beta_1')g(\beta_0')}{\iint g(\boldsymbol{Z}|\beta_1', \beta_0')g(\beta_1')g(\beta_0')\, d\beta_1'\, d\beta_0'} \\
&\propto g(\boldsymbol{Z}|\beta_1', \beta_0')g(\beta_1')g(\beta_0') \\
&= \left\{\prod_{l=0}^{1}(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}\left\{\sqrt{\frac{\tau_{\beta_1'}}{2\pi}}e^{-\frac{\tau_{\beta_1'}}{2}(\beta_1'-\mu_{\beta_1'})^2}\right\} \\
&\left\{\sqrt{\frac{\tau_{\beta_0'}}{2\pi}}e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0'-\mu_{\beta_0'})^2}\right\} \\
&\propto \left\{\prod_{l=0}^{1}(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}\left\{e^{-\frac{\tau_{\beta_1'}}{2}(\beta_1'-\mu_{\beta_1'})^2}\right\} \\
&\left\{e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0'-\mu_{\beta_0'})^2}\right\},
\end{aligned}
$$

where $\boldsymbol{Z} = (Z_0, Z_1)$. So the full conditionals of β_1' and β_0' are

$$
\begin{aligned}
&g(\beta_1'|\boldsymbol{D}_0, \beta_0') = \frac{g(\beta_1', \beta_0'|\boldsymbol{D}_0)}{g(\beta_0'|\boldsymbol{D}_0)} \\
&= \frac{g(\beta_1', \beta_0'|\boldsymbol{D}_0)}{\int g(\beta_1', \beta_0'|\boldsymbol{D}_0)\, d\beta_1'} \\
&= \frac{\left\{\prod_{l=0}^{1}(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}\left\{e^{-\frac{\tau_{\beta_1'}}{2}(\beta_1'-\mu_{\beta_1'})^2}\right\}}{\int\left\{\prod_{l=0}^{1}(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}\left\{e^{-\frac{\tau_{\beta_1'}}{2}(\beta_1'-\mu_{\beta_1'})^2}\right\}\, d\beta_1'} \\
&\propto \left\{\prod_{l=0}^{1}(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}\left\{e^{-\frac{\tau_{\beta_1'}}{2}(\beta_1'-\mu_{\beta_1'})^2}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&g(\beta_0'|\boldsymbol{D}_0, \beta_1') = \frac{g(\beta_1', \beta_0'|\boldsymbol{D}_0)}{g(\beta_1'|\boldsymbol{D}_0)} \\
&= \frac{g(\beta_1', \beta_0'|\boldsymbol{D}_0)}{\int g(\beta_1', \beta_0'|\boldsymbol{D}_0)\, d\beta_0'} \\
&= \frac{\left\{\prod_{l=0}^{1}(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}\left\{e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0'-\mu_{\beta_0'})^2}\right\}}{\int\left\{\prod_{l=0}^{1}(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}\left\{e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0'-\mu_{\beta_0'})^2}\right\}\, d\beta_0'} \\
&\propto \left\{\prod_{l=0}^{1}\{\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}}\}^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}\left\{e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0'-\mu_{\beta_0'})^2}\right\}.
\end{aligned}
$$

Finally, we do power prior analysis (2.12) with both data $\boldsymbol{D}$ and $\boldsymbol{D}_0$ to draw posterior samples of β_1 from $h(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0, a)$. The augmented posterior of β_1,β_0,$\boldsymbol{b}$,τ_b,β_1',β_0',ρ_1, μ_1,τ_1 is

$$
\begin{aligned}
&h(\beta_1, \beta_0, \boldsymbol{b}, \tau_b, \beta_1', \beta_0', \mu_1, \tau_1, \rho_1|\boldsymbol{D}, \boldsymbol{D}_0, a) \\
&= \frac{h(\beta_1, \beta_0, \boldsymbol{b}, \tau_b, \beta_1', \beta_0', \mu_1, \tau_1, \rho_1, \boldsymbol{D}, \boldsymbol{D}_0, a)}{h(\boldsymbol{D}, \boldsymbol{D}_0, a)} \\
&\propto L(\beta_1, \beta_0|\boldsymbol{D}, \boldsymbol{b}) f(\boldsymbol{b}|\tau_b) f(\tau_b) f(\beta_0) \{L(\beta_1', \beta_0'|\boldsymbol{D}_0)\}^a g(\beta_0') h(\boldsymbol{\beta_1}|\boldsymbol{\mu_1}, \boldsymbol{\Lambda_1}) \\
&h_1(\mu_1|a_1, R_1) h_1(\tau_1|\kappa_1, \nu_1) h_1(\rho_1|c_1, d_1) \\
&= \left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{\prod_{i=1}^{I}\sqrt{\frac{\tau_b}{2\pi}}e^{-\frac{1}{2}b_i^2\tau_b}\right\} \\
&\left\{\frac{\nu_b^{\kappa_b}}{\Gamma(\kappa_b)}\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}\left\{\sqrt{\frac{\tau_{\beta_0}}{2\pi}}e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\} \\
&\left\{\prod_{l=0}^{1}(\frac{e^{\beta_0'+\beta_1' X_{1l}'}}{1+e^{\beta_0'+\beta_1' X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1' X_{1l}'}}{1+e^{\beta_0'+\beta_1' X_{1l}'}})^{N_l-Z_l}\right\}^a \\
&\left\{\sqrt{\frac{\tau_{\beta_0'}}{2\pi}}e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0'-\mu_{\beta_0'})^2}\right\}\left\{\frac{1}{2\pi}|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{\sqrt{\frac{R_1}{2\pi}}e^{-\frac{R_1}{2}(\mu_1-a_1)^2}\right\} \\
&\left\{\frac{\nu_1^{\kappa_1}}{\Gamma(\kappa_1)}\tau_1^{\kappa_1-1}e^{-\nu_1\tau_1}\right\}\left\{\frac{\rho_1^{c_1-1}(1-\rho_1)^{d_1-1}}{B(c_1,d_1)}\right\} \\
&\propto \left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1 X_{1i}+b_i}}{1+e^{\beta_0+\beta_1 X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\} \\
&\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\}\left\{\prod_{l=0}^{1}(\frac{e^{\beta_0'+\beta_1' X_{1l}'}}{1+e^{\beta_0'+\beta_1' X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1' X_{1l}'}}{1+e^{\beta_0'+\beta_1' X_{1l}'}})^{N_l-Z_l}\right\}^a \\
&\left\{e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0'-\mu_{\beta_0'})^2}\right\}\left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{e^{-\frac{R_1}{2}(\mu_1-a_1)^2}\right\} \\
&\left\{\tau_1^{\kappa_1-1}e^{-\nu_1\tau_1}\right\}\left\{\rho_1^{c_1-1}(1-\rho_1)^{d_1-1}\right\},
\end{aligned}
$$

so the full conditionals of β_1,β_0,$\boldsymbol{b}$,τ_b,β_1', β_0',ρ_1, μ_1 and τ_1 are

$$
\begin{aligned}
&h(\beta_1|\boldsymbol{D},\boldsymbol{D}_0,a,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1)\\
&=h(\beta_1|\boldsymbol{D},\beta_0,\boldsymbol{b},\beta_1',\mu_1,\tau_1,\rho_1)\\
&=\frac{h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\beta_1}\\
&=\frac{\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{|\boldsymbol{\Lambda_h}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}}{\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{|\boldsymbol{\Lambda_h}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}d\beta_1}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{|\boldsymbol{\Lambda_h}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\\
&\left\{e^{-\frac{\tau_1}{2(1-\rho_1^2)}\{(\beta_1-\mu_1)^2-2\rho_1(\beta_1-\mu_1)(\beta_1'-\mu_1)\}}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&h(\beta_0|\boldsymbol{D},\boldsymbol{D}_0,a,\beta_1,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1)\\
&=h(\beta_0|\boldsymbol{D},\beta_1,\boldsymbol{b})\\
&=\frac{h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\beta_0}\\
&=\frac{\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\}}{\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\}d\beta_0}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&h(\boldsymbol{b}|\boldsymbol{D},\boldsymbol{D}_0,a,\beta_1,\beta_0,\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1)\\
&=h(\boldsymbol{b}|\boldsymbol{D},\beta_0,\beta_1,\tau_b)\\
&=\frac{h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int\cdots\int h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\boldsymbol{b}}\\
&=\frac{\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}}{\int\cdots\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}d\boldsymbol{b}}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}(\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{Y_{ij}}(1-\frac{e^{\beta_0+\beta_1X_{1i}+b_i}}{1+e^{\beta_0+\beta_1X_{1i}+b_i}})^{1-Y_{ij}}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&h(\tau_b|\boldsymbol{D},\boldsymbol{D}_0,a,\beta_1,\beta_0,\boldsymbol{b},\beta_1',\beta_0',\mu_1,\tau_1,\rho_1)\\
&=h(\tau_b|\boldsymbol{b})\\
&=\frac{h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\tau_b}\\
&=\frac{\left\{\prod_{i=1}^I\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}}{\int\left\{\prod_{i=1}^I\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}\,d\tau_b}\\
&\propto\left\{\prod_{i=1}^I\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}\\
&\sim\text{Gamma}(\kappa_b+\frac{I}{2},\nu_b+\frac{\sum_{i=1}^I b_i^2}{2}),
\end{aligned}
$$

$$
\begin{aligned}
&h(\beta_1'|\boldsymbol{D},\boldsymbol{D}_0,a,\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_0',\mu_1,\tau_1,\rho_1)\\
&=h(\beta_1'|\boldsymbol{D}_0,a,\beta_0',\beta_1,\mu_1,\tau_1,\rho_1)\\
&=\frac{h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\beta_1'}\\
&=\frac{\left\{\prod_{l=0}^1(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}^a\left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}}{\int\left\{\prod_{l=0}^1(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}^a\left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\,d\beta_1'}\\
&\propto\left\{\prod_{l=0}^1(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}^a\left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\\
&\propto\left\{\prod_{l=0}^1(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}^a\left\{e^{-\frac{\tau_1}{2(1-\rho_1^2)}\{(\beta_1'-\mu_1)^2-2\rho_1(\beta_1-\mu_1)(\beta_1'-\mu_1)\}}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&h(\beta_0'|\boldsymbol{D},\boldsymbol{D}_0,a,\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\mu_1,\tau_1,\rho_1)\\
&=h(\beta_0'|\boldsymbol{D}_0,a,\beta_1')\\
&=\frac{h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\beta_0'}\\
&=\frac{\left\{\prod_{l=0}^1(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}^a e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0'-\mu_{\beta_0'})^2}}{\int\left\{\prod_{l=0}^1(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}^a e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0'-\mu_{\beta_0'})^2}\,d\beta_0'}\\
&\propto\left\{\prod_{l=0}^1(\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{Z_l}(1-\frac{e^{\beta_0'+\beta_1'X_{1l}'}}{1+e^{\beta_0'+\beta_1'X_{1l}'}})^{N_l-Z_l}\right\}^a e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0'-\mu_{\beta_0'})^2},
\end{aligned}
$$

$$
\begin{aligned}
&h(\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a,\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1)\\
&= h(\rho_1|\beta_1,\beta_1',\mu_1,\tau_1)\\
&= \frac{h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\rho_1}\\
&= \frac{\left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{\rho_1^{c_1-1}(1-\rho_1)^{d_1-1}\right\}}{\int\left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{\rho_1^{c_1-1}(1-\rho_1)^{d_1-1}\right\}\,d\rho_1}\\
&\propto \left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{\rho_1^{c_1-1}(1-\rho_1)^{d_1-1}\right\}\\
&\propto \rho_1^{c_1-1}(1-\rho_1)^{d_1-1}e^{-\tau_1\frac{(\beta_1-\mu_1)^2-2\rho_1(\beta_1-\mu_1)(\beta_1'-\mu_1)+(\beta_1'-\mu_1)^2}{2(1-\rho_1^2)}}.
\end{aligned}
$$

$$
\begin{aligned}
&h(\mu_1|\boldsymbol{D},\boldsymbol{D}_0,a,\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\tau_1,\rho_1)\\
&= h(\mu_1|\beta_1,\beta_1',\tau_1,\rho_1)\\
&= \frac{h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\beta_1,\beta_0,\boldsymbol{b},\tau_b,\beta_1',\beta_0',\mu_1,\tau_1,\rho_1|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\mu_1}\\
&= \frac{\left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{e^{-\frac{R_1}{2}(\mu_1-a_1)^2}\right\}}{\int\left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{e^{-\frac{R_1}{2}(\mu_1-a_1)^2}\right\}\,d\mu_1}\\
&\propto \left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{e^{-\frac{R_1}{2}(\mu_1-a_1)^2}\right\}\\
&\propto e^{-\frac{1}{2}\{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})+R_1(\mu_1-a_1)^2\}}\\
&\propto e^{-\frac{1}{2}(\frac{2\tau_1}{1+\rho_1}+R_1)(\mu_1-\frac{\frac{\tau_1}{1+\rho_1}(\beta_1+\beta_1')+R_1a_1}{\frac{2\tau_1}{1+\rho_1}+R_1})^2}\\
&\sim \mathrm{N}(\frac{\frac{\tau_1}{1+\rho_1}(\beta_1+\beta_1')+R_1a_1}{\frac{2\tau_1}{1+\rho_1}+R_1},\frac{2\tau_1}{1+\rho_1}+R_1),
\end{aligned}
$$

$$
\begin{aligned}
&h(\tau_1|\boldsymbol{D}, \boldsymbol{D}_0, a, \beta_1, \beta_0, \boldsymbol{b}, \tau_b, \beta_1', \beta_0', \mu_1, \rho_1) \\
&= h(\tau_1|\beta_1, \beta_1', \mu_1, \rho_1) \\
&= \frac{h(\beta_1, \beta_0, \boldsymbol{b}, \tau_b, \beta_1', \beta_0', \mu_1, \tau_1, \rho_1|\boldsymbol{D}, \boldsymbol{D}_0, a)}{\int h(\beta_1, \beta_0, \boldsymbol{b}, \tau_b, \beta_1', \beta_0', \mu_1, \tau_1, \rho_1|\boldsymbol{D}, \boldsymbol{D}_0, a)\, d\tau_1} \\
&= \frac{\left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{\tau_1^{\kappa_1-1} e^{-\nu_1\tau_1}\right\}}{\int \left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{\tau_1^{\kappa_1-1} e^{-\nu_1\tau_1}\right\} d\tau_1} \\
&\propto \left\{|\boldsymbol{\Lambda_1}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_1}-\boldsymbol{m_1})'\boldsymbol{\Lambda_1}(\boldsymbol{\beta_1}-\boldsymbol{m_1})}{2}}\right\}\left\{\tau_1^{\kappa_1-1} e^{-\nu_1\tau_1}\right\} \\
&= \tau_1^{\kappa_1+\frac{1}{2}-1} e^{-\tau_1(\nu_1+\frac{(\beta_1-\mu_1)^2-2\rho_1(\beta_1-\mu_1)(\beta_1'-\mu_1)+(\beta_1'-\mu_1)^2}{2(1-\rho_1^2)})} \\
&\sim \text{Gamma}(\kappa_1+\frac{1}{2}, \nu_1+\frac{(\beta_1-\mu_1)^2-2\rho_1(\beta_1-\mu_1)(\beta_1'-\mu_1)+(\beta_1'-\mu_1)^2}{2(1-\rho_1^2)}),
\end{aligned}
$$

The Metropolis Hastings Component of the Hybrid Algorithm

When Gibbs algorithm is used to draw posterior samples of β_1 from $f(\beta_1|\boldsymbol{D})$, β_1' from $g(\beta_1'|\boldsymbol{D}_0)$ or β_1 from $h(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0, a)$, some of the full conditionals do not have closed forms. So we construct a markov chain with the full conditional as target distribution to get the corresponding posterior samples. This is the MH part of the hybrid algorithm. We develop a random walk MH algorithm with a t proposal density. Since the algorithm can be implemented for the parameters in the same way, we write the MH algorithm in a more general notation.

We write the parameter of interest and the nuisance parameters as $\boldsymbol{\theta}$ and $\boldsymbol{\theta'}$, respectively, and write the data by $\boldsymbol{D}_g$. Assuming the full conditional is $k(\boldsymbol{\theta}|\boldsymbol{D}_g, \boldsymbol{\theta'})$, we note that the MH part of the hybrid algorithm for $\boldsymbol{\theta}$ can be carried out as follows.

1. Start with $\boldsymbol{\theta}^0$ and set $m = 1$.

2. Sample $\boldsymbol{\theta}^* \sim t(\boldsymbol{\theta}^{m-1}, \nu_t)$, and accept $\boldsymbol{\theta}^*$ with probability

$$\min\{1, \frac{k(\boldsymbol{\theta}^*|\boldsymbol{D}_g, \boldsymbol{\theta'})}{k(\boldsymbol{\theta}^{m-1}|\boldsymbol{D}_g, \boldsymbol{\theta'})}\},$$

and set $\boldsymbol{\theta}^m = \boldsymbol{\theta}^*$. Otherwise, set $\boldsymbol{\theta}^m = \boldsymbol{\theta}^{m-1}$.

3. Increase m by 1 and return to step 2.

Here $t(\boldsymbol{\theta}^{m-1}, \nu_t)$ is the univariate or multivariate t distribution with $\boldsymbol{\theta}^{m-1}$ and ν_t as the location and degree of freedom, depending on whether $\boldsymbol{\theta}$ is a random variable or vector. With this hybrid algorithm, the posterior samples of $\boldsymbol{\theta}$ are drawn from its posterior $k(\boldsymbol{\theta}|\boldsymbol{D}_g)$.

2.5 Simulation Studies

2.5.1 Simulation Settings and Data Generation

A simulation study was carried out to examine the performance of the power prior method in a two-arm cluster randomized trial with a binary outcome. Without loss of generality, we assumed both current and historical trials were balanced in treatment group size. The model for generating the current data $\boldsymbol{D}$ was $Y_{ij} \sim \text{Bernoulli}(p_i)$, $\text{logit}(p_i) = \beta_0 + \beta_1 X_{1i} + b_i, i = 1, 2, \ldots, I, j = 1, 2, \ldots, J$, $b_i \sim \text{N}(0, \tau_b)$, and the model for generating historical data $\boldsymbol{D}_0$ was $Z_l \sim \text{Binomial}(N, q_l)$, $\text{logit}(q_l) = \beta_0' + \beta_1' X_{1l}', l = 0, 1$. In all simulated datasets, we assumed the first half of X_{1i} were 0, and the remaining were 1. $X_{1l}' = 0$ if $l = 0$ and 1 otherwise.

Five different scenarios were considered, which were designed to investigate the effect of (1) the treatment effect (β_1, β_1'); (2) the historical sample size N; (3) the total cluster number I; (4) the cluster size J; and (5) the intra cluster correlation

coefficient (ICC) ρ in current trial, where $\rho = \frac{1}{1+\frac{\pi^2}{3}\tau_b}$. We considered four unique precision values $\tau_b = 0.199, 0.463, 1, 2.687$, which corresponded to the ICC values $\rho = 0.604, 0.396, 0.233, 0.102$. The settings of current and historical trials were given in Tables $2.1 - 2.2$, respectively.

The two treatment effects were assumed to be exchangeable in the power prior method, and this assumption implied that β_1 and β_1' were random variables. Conceptually, the joint distribution of β_1 and β_1' should be specified to generate multiple replicates of (β_1,β_1') for use. However, this strategy was too computationally extensive to be feasible in practice. Instead, we could just consider a few typical values of (β_1, β_1'), and it could depict the profile of the effect of (β_1, β_1').

In each scenario, $1,000$ simulated datasets were generated, and for each simulated dataset, the power prior method was applied with four different discounting fractions, i.e. 0, 1, a^{sym} and a^{asym}. The posterior median, standard deviation (SD) and the 95% highest posterior density (HPD) interval of β_1 were computed. Furthermore, the 95% HPD interval was used to test the null hypothesis $\text{H}_0 : \text{E}(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0) = 0$. With the $1,000$ simulated datasets, the average of the posterior medians, the average of the posterior SDs, the relative bias, the root of the mean squared error (rMSE) as well as the average length, the coverage probability (CP) and rejection probability (RP) of the 95% HPD interval were calculated. The RP was just the type I error rate when the true value of β_1 was 0 or power otherwise.

Given a specific scenario, the simulation study was carried out as follows.

1. Generate current data $\boldsymbol{D}$ with $(I, J, \beta_0, \beta_1, \tau_b)$ from Table 2.1 and generate historical data $\boldsymbol{D}_0$ with (N, β_0', β_1') from Table 2.2.

2. Fit current data $\boldsymbol{D}$ with Model (2.4) and get the posterior samples of β_1 from $f(\beta_1|\boldsymbol{D})$ with the MH within Gibbs algorithm, and fit historical data $\boldsymbol{D}_0$ with Model (2.5) and get the posterior samples of β_1' from $g(\beta_1'|\boldsymbol{D}_0)$ with the same algorithm.
3. Calculate the symmetric and asymmetric KL divergences with k-NN algorithm by using these two posterior samples, and take them as the symmetric and asymmetric discounting fractions after the inverse exponential transformation.
4. Do the power prior analysis under four different discounting fractions, i.e. 0, 1, a^{sym} and a^{asym}. In each case, get a posterior sample of β_1 from $h(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0, a)$ with the MH within Gibbs algorithm.
5. Calculate the posterior median, SD and 95% HPD interval with the posterior samples from $h(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0, a)$, and calculate the 95% HPD interval with the posterior samples from $h(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0, a)$ and use it to determine (1) its length L, (2) whether it covers the true value of β_1, and let $C = 1$ if true and 0 otherwise, (3) and whether it covers 0, and let $R = 1$ if true and 0 otherwise.
6. Repeat the above steps $1,000$ times, and report the average of the discounting fractions a^{sym} and a^{asym} and the average of the posterior median, the average of the posterior SD, the relative bias and root of the mean squared error (rMSE). In addition, report the average of L, C and R as the average length, CP and RP of the 95% HPD interval.

In the MH within Gibbs algorithm, one Markov Chain of length $2,500$ was run with the first $1,500$ samples discarded and the thinning rate being 5 in the separate data analysis of current and historical trials under non-informative prior, while one Markov Chain of length $5,000$ was run with the first $2,500$ samples discarded and the

thinning rate being 5 in the power prior analysis. In both cases, the degree of freedom in the t proposal density, ν_t, was set to be 3. The MCMC convergence was diagnosed with the trace plot. When fitting the current data alone with model (2.4), we specified non-informative uniform prior for τ_b instead of non-informative gamma prior because the former is more robust in hierarchical models (Browne et al., 2006; Gelman et al., 2006). In the k-NN algorithm, the maximum number of nearest neighbors to search, k, was set to be 100.

2.5.2 Simulation Results

For illustrative purpose, here we only reported the discounting fraction estimate, the posterior SD, the rMSE of the posterior median and the RP of the 95% HPD interval for (1) $\beta_1 = \log(1) = 0$ in scenario 1; (2) $\beta_1 = \log(4) \approx 1.386$ in scenario 1; (3) $N = 1,000$ in scenario 2; (4) $I = 10$ in scenario 3; (5) $J = 10$ in scenario 4; and (6) $\tau_b = 0.463$ in scenario 5, i.e. $\rho = 0.396$. All these results are summarized in Figures $2.1 - 2.6$. The other results could be obtained from the authors once requested.

Figures 2.1 and 2.2 showed that a^{sym} and a^{asym} would increase as β_1' got closer to β_1, but they would never attach 1 even when β_1 and β_1' were equal. This was because a^{sym} and a^{asym} measured the homogeneity of the two sources of data rather than the similarity between the two treatment effects. Secondly, as a^{sym} and a^{asym} became larger, more percentage of historical information would be incorporated, and as a result, the posterior SD and rMSE became smaller and the type I error rate and power would increase. However, the maximum value of power did not achieve at the maximum value of the discounting fraction. This was because the total amount of historical information was not largest at that time.

Compare Figures $2.3 - 2.6$ to Figure 2.2, we could study the effects of N, I, J and ρ on a^{sym}, a^{asym}, SD, rMSE and power. a^{sym} and a^{asym} would get smaller as N and ρ increased and I decreased, but J seemed to have little effect on these two discounting fractions. This was a good property because it meant the power prior procedure did not allow the historical information to dominate current data analysis. In addition, SD and rMSE became smaller and power became higher as I and J increased and ρ decreased, but they were not affected greatly by N. Therefore, given a current trial, the use of a more similar rather than larger historical trial played a key role in improving the precision of the estimate and boosting the power.

Although Figures $2.2 - 2.6$ demonstrated that the power prior method could increase the power compared to not borrowing any historical information when historical treatment lay in certain range of current one, Figure 2.1 showed that the type I error rate was inflated at the same time. Therefore, it was hard to see how much the auxiliary information contributed to the net gain in power. Though it was not possible to control the type I error rate at the same value across different methods for comparing the gain in power in simulation studies, Figures $2.1 - 2.6$ still delivered some useful information. For example, in Figure 2.2, when $\beta_1' = \log(5)$, asymmetric discounting fraction resulted in a larger power than symmetric one, and Figure 2.1 showed that asymmetric discounting fraction led to a smaller type I error rate at the same point. It meant that asymmetric discounting fraction had a better performance in inference than symmetric one in that case. However, this might be just a local property and could not be straightforwardly extended to its neighborhood because the power function was not necessarily smooth.

2.6 Example

2.6.1 Data Source

To illustrate, we applied the proposed method in the analysis of a two-arm cluster randomized trial. The trial was sponsored by Merck Pharmaceutical and it was designed to assess the efficacy of an electronic vaccination reminder intervention, aimed at improving the first dosing uptake of a quadrivalent human papillomavirus (HPV) vaccine in children and adolescents. The trial was conducted in five pediatric clinics, started in Feb. 2015. All pediatricians and nurse practitioners who provided health care at the participating clinics were included in the trial. Randomization was performed at the physician level. Specifically, 30 physicians were randomly assigned to one of the two arms. Nineteen physicians were assigned to the intervention arm, and 11 in the control arm. In the intervention arm, immunization reminders were delivered to the physician through an electronic medical record system. The system determined patient's eligibility of receiving the HPV vaccine. Physicians of eligible patients received reminders when patient was present in the clinic. No reminders were provided for patients of physicians in the control group, although the control physicians retained the rights to order vaccination as they saw fit. Patients eligible for HPV vaccination under the care of intervention or control physicians were recruited for study participation. Within one year of the study period, a total of 475 eligible patients were recruited. Numbers of patients seen by each physician (i.e., the cluster size) varied by physician, ranged from 1 to 55.

There is a sizable literature on vaccine reminder interventions, with different scripts and contents, for different clinical populations. Modes of message delivery

also varied. For illustration, we selected a simple randomized trial conducted by Szilagyi et al. (2011) between Oct. 2007 and Dec. 2008 as the historical trial. As in the current trial, this historical trial was also a two-arm trial on the efficacy of one reminder intervention on the uptake rate of the first dose of the HPV vaccine. The intervention messages were also delivered to the patients. The trial did not include male children. Briefly, 1055 and 1084 eligible girls were recruited and randomized to the control and intervention arm, respectively, and at the end of the study, 453 and 634 girls received the first dose of HPV vaccine in two treatment arms.

2.6.2 Analytical Results

We analyzed the historical data with Model (2.5) and performed power prior analysis (2.12) to the current data by using the historical data under four discounting fractions, i.e. 0,1,a^{sym} and a^{asym}. The power prior analysis under the discounting fraction 0 was equivalent to analyze the current data with Model (2.4). All analytic results were summarized in Table 2.3, and they were labeled Analysis $1-5$ accordingly.

The results in Table 2.3 demonstrated that incorporation of historical information into current data analysis could reduce the standard deviation and the length of the 95% HPD interval compared to not borrowing any historical evidence, and the more historical information the power prior procedure borrowed, the larger the reduction was. However, the precision of the treatment effect estimate yielded by the power prior procedure was higher than that by historical data analysis alone. It meant that the amount of current information plus the discounted historical information was less than the total historical information. Furthermore, the treatment effect estimate from the power prior procedure lay between those from separate data analysis of both

trials, and it implied that the power prior procedure would shrink the current treatment effect towards the historical one. Separate data analysis results showed that the 95% HPD interval of the current treatment effect covered 0 and that of the historical treatment effect did not, which induced the conclusion that the current treatment effect was not significant but the historical treatment effect was. When borrowing historical information through power prior under the symmetric discounting fraction, the same conclusion about the current treatment effect was made though the 95% HPD interval became narrower. When asymmetric discounting fraction was adopted, more historical strength was borrowed, then the opposite conclusion was made, that is, the current treatment effect was significant.

2.7 Discussion

In this chapter, a Bayesian power prior method was developed to analyze the treatment effect in a two-arm cluster randomized trial of binary outcome. A novel power prior was constructed under the assumption that the treatment effects in the two trials were exchangeable, and a new discounting fraction function was proposed as the inverse exponential transformation of the Kullback Leibler divergence of the posterior distributions of the two treatment effects, which were derived from Bayesian data analysis of the two trials under non-informative priors separately. Through the power prior, the historical information of the treatment effect could be appropriately incorporated into the current data analysis, thereby decreasing the variability of the treatment effect estimate as well as boosting the power with non-significant inflation of type I error rate in certain situations compared to not borrowing external evidence.

The inverse exponential transformation was adopted in the construction of the discounting fraction function, and therefore the power prior was identified with this discounting fraction function. Such a power prior could help increase the power, but it would also slightly inflate the type I error rate. This meant the inverse exponential transformation might not be the optimal one. Future work can be done to find out another transformation function that results in a better trade off between type I error rate inflation and power increase. Another limit is that this power prior is not compared to other types of power priors (Ibrahim et al., 2015). Such a comparison would help demonstrate the merits and drawbacks of this power prior.

Additionally, only one historical trial was considered in this chapter, which limited the use of this procedure in application because more than one historical trials usually existed. Though a straightforward way to extend this method to the case of multiple historical trials was mentioned in Equation (2.2), it brings cumbersome and extensive computations when the number of historical trials available is large, which makes such a simple extension less attractive. One possible way to bypass this dilemma is to pool the historical data first and then construct the power prior for the meta data. But before taking this strategy, we need to prove its equivalence to the first simple extension way. In the end, the power prior method provides a potential to design a new two-arm cluster randomized trial of binary outcome by using relevant historical data. One promising advantage of such a design strategy is that fewer samples may be needed to achieve certain power compared to not borrowing any external information. But since no explicit formula for sample size exists, a numerical algorithm must be developed to address the sample size determination problem.

Table 2.1: Setting of two-arm current trial with binary outcome

Scenario	I	J	β_0	β_1	τ_b
1	20	20	-0.1	$\log(1), \log(2), \ldots, \log(6)$	1
2	20	20	-0.1	$\log(4)$	1
3	$10, 20, 40$	20	-0.1	$\log(4)$	1
4	20	$10, 20, 40$	-0.1	$\log(4)$	1
5	20	20	-0.1	$\log(4)$	$0.199, 0.463, 1, 2.687$

Table 2.2: Setting of two-arm historical trial with binary outcome

Scenario	N	β_0'	β_1'
1	200	0.1	$\log(1), \log(2), \ldots, \log(6)$
2	$100, 200, 400, 800, 1000$	0.1	$\log(1), \log(2), \ldots, \log(6)$
3	200	0.1	$\log(1), \log(2), \ldots, \log(6)$
4	200	0.1	$\log(1), \log(2), \ldots, \log(6)$
5	200	0.1	$\log(1), \log(2), \ldots, \log(6)$

Table 2.3: Estimates of discounting fraction and intervention effect

Analysis	Discounting fraction	Posterior median	Posterior SD	95% HPD interval
1		0.628	0.087	$[0.458, 0.798]$
2	0	0.808	0.452	$[-0.122, 1.659]$
3	1	0.749	0.379	$[0.047, 1.601]$
4	0.075	0.759	0.424	$[-0.091, 1.612]$
5	0.262	0.747	0.405	$[0.013, 1.638]$

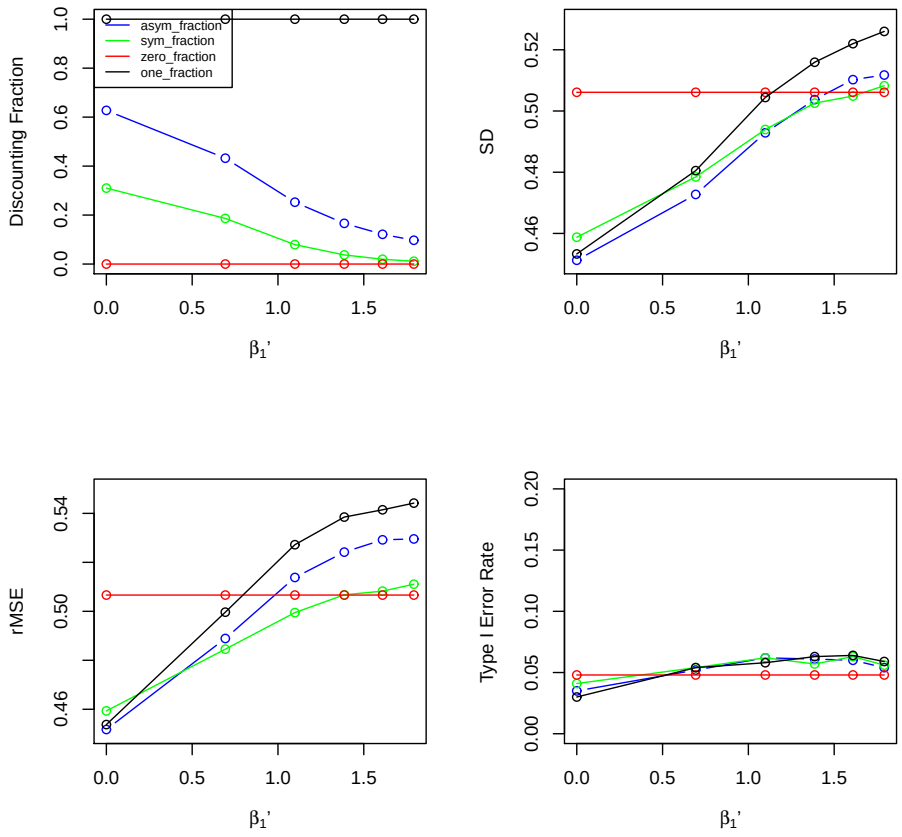

Figure 2.1: In two-arm trials with binary outcome, the effect of the historical treatment effect β_1' on the discounting fraction, SD, rMSE and type I error rate where the treatment effect $\beta_1 = \log(1) = 0$, the total number of cluster $I = 20$, the cluster size $J = 20$ and the ICC $\rho = 0.233$ in current trial and the arm size $N = 200$ in historical trial

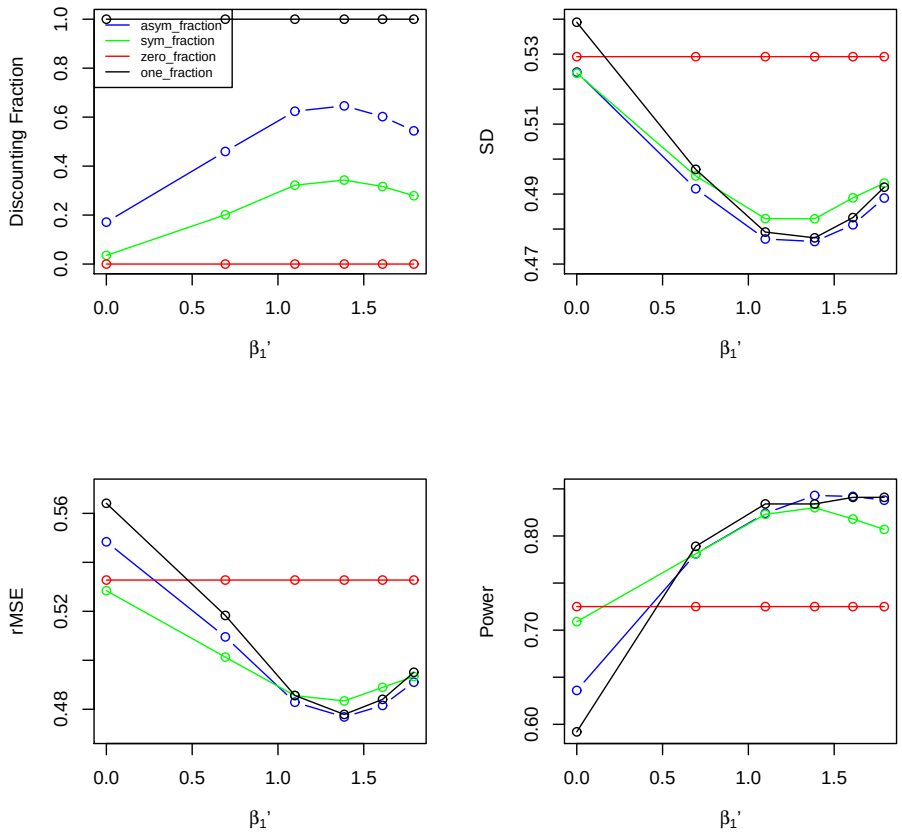

Figure 2.2: In two-arm trials with binary outcome, the effect of the historical treatment effect β_1' on the discounting fraction, SD, rMSE and power where the treatment effect $\beta_1 = \log(4) \approx 1.386$, the total number of cluster $I = 20$, the cluster size $J = 20$ and the ICC $\rho = 0.233$ in current trial and the arm size $N = 200$ in historical trial

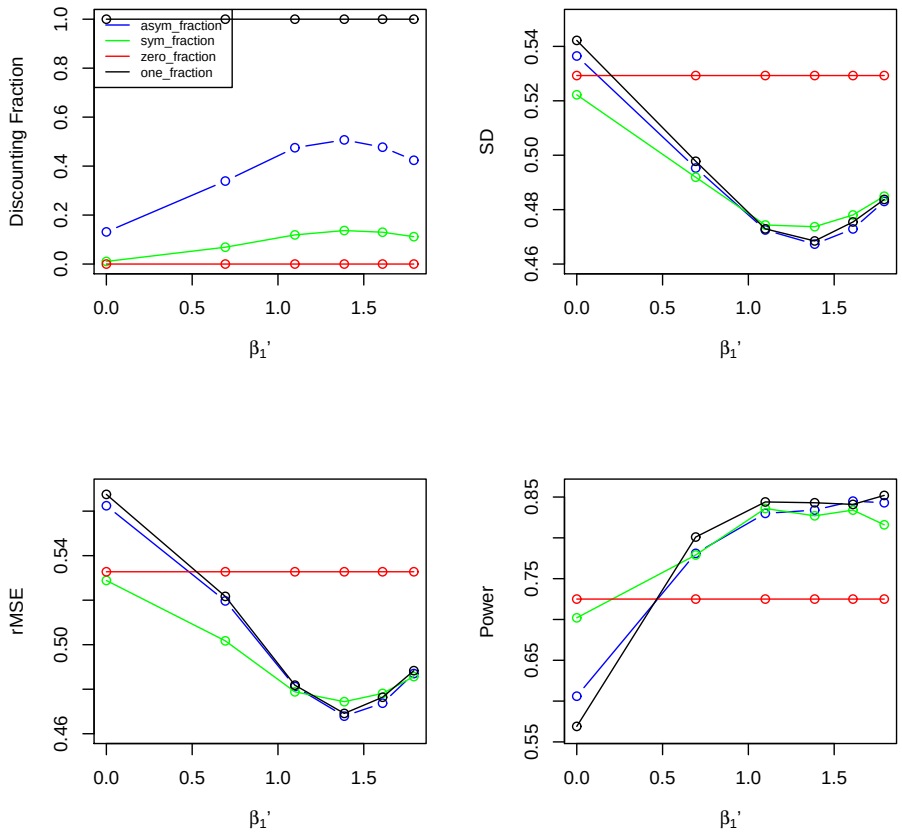

Figure 2.3: In two-arm trials with binary outcome, the effect of the historical treatment effect β_1' on the discounting fraction, SD, rMSE and power where the treatment effect $\beta_1 = \log(4) \approx 1.386$, the total number of cluster $I = 20$, the cluster size $J = 20$ and the ICC $\rho = 0.233$ in current trial and the arm size $N = 1000$ in historical trial

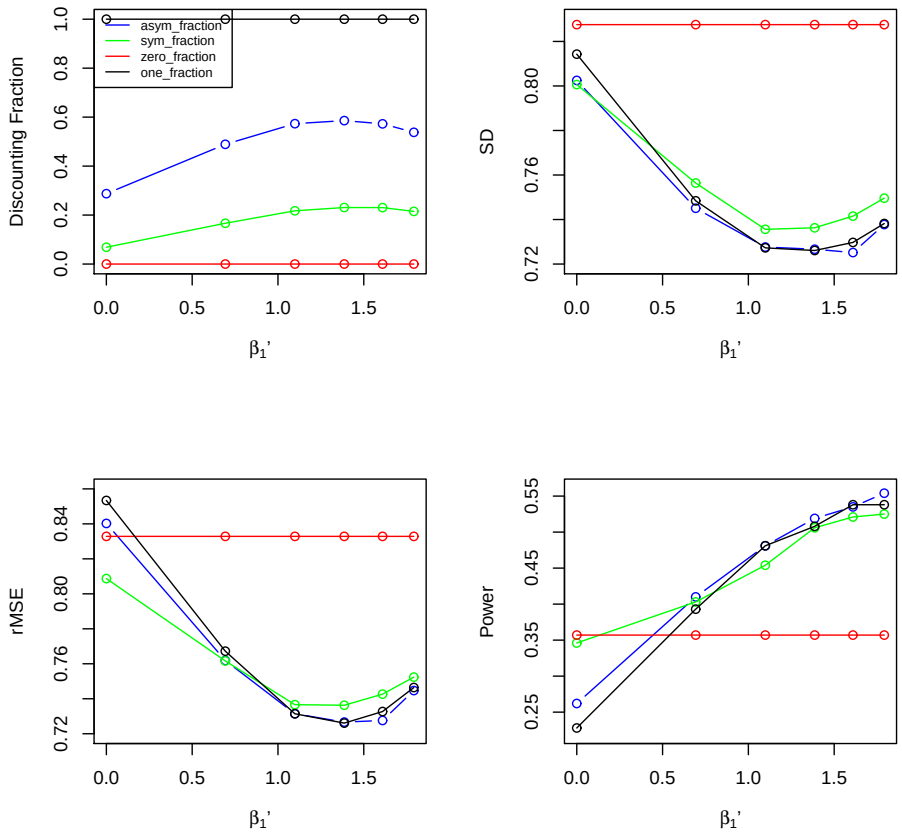

Figure 2.4: In two-arm trials with binary outcome, the effect of the historical treatment effect β_1' on the discounting fraction, SD, rMSE and power where the treatment effect $\beta_1 = \log(4) \approx 1.386$, the total number of cluster $I = 10$, the cluster size $J = 20$ and the ICC $\rho = 0.233$ in current trial and the arm size $N = 200$ in historical trial

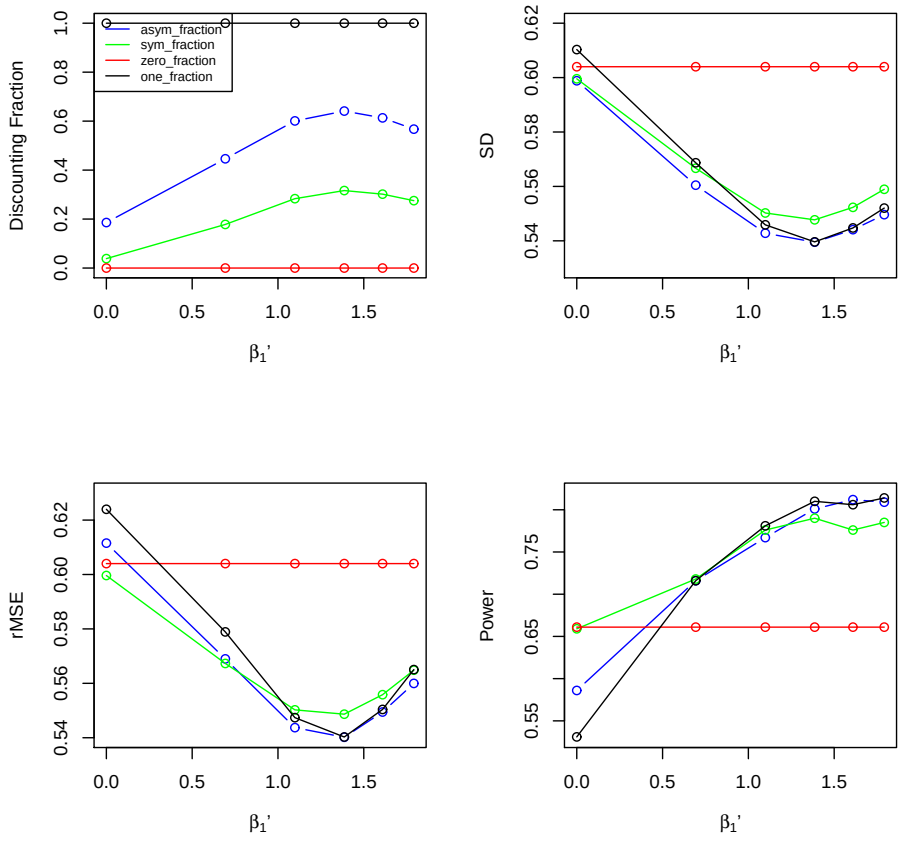

Figure 2.5: In two-arm trials with binary outcome, the effect of the historical treatment effect β_1' on the discounting fraction, SD, rMSE and power where the treatment effect $\beta_1 = \log(4) \approx 1.386$, the total number of cluster $I = 20$, the cluster size $J = 10$ and the ICC $\rho = 0.233$ in current trial and the arm size $N = 200$ in historical trial

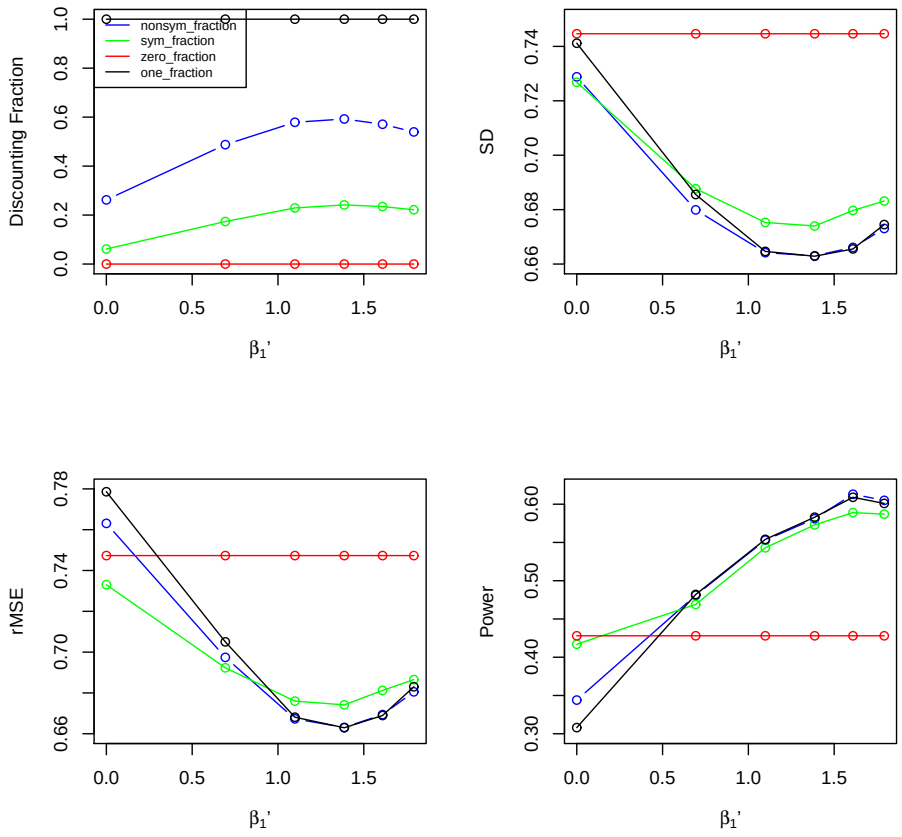

Figure 2.6: In two-arm trials with binary outcome, the effect of the historical treatment effect β_1' on the discounting fraction, SD, rMSE and power where the treatment effect $\beta_1 = \log(4) \approx 1.386$, the total number of cluster $I = 10$, the cluster size $J = 20$ and the ICC $\rho = 0.396$ in current trial and the arm size $N = 200$ in historical trial

Chapter 3

Design of Two-arm Cluster Randomized Trials with Binary Outcome

3.1 Cluster Randomized Trial Design

Sample size determination is an important task in the design of clinical trials. For cluster randomized trials, sample size calculation is complicated because the randomization at the cluster level creates an interdependency among the subjects within the same cluster. As a result, and the power of the analysis depends not only on the numbers of clusters and patients, but also the strength of intra-cluster correlation (ICC) (Donner, 1992; Donner and Klar, 1994; Kerry and Bland, 1998). A sample size formula for binary outcomes is presented by Donner and Klar (2000).

For all practical purposes, the true values of ICC are often unknown, so they have to be estimated based either on expert opinions or on pilot data. In some situations, the maximum value of ICC is used to ensure a conservative power estimate (Rotondi and Donner, 2009). Sensitivity analysis is often performed to evaluate the influences of different ICC values (Turner et al., 2005).

Further complicating the power analysis is the unequal cluster size. In the case of varying cluster sizes, analysts often use the average size of the clusters in the standard sample size formulae. While the method is simple to use, its accuracy is often less than certain. In some situations, the cluster size is assumed to be a random variable, and new formulae are developed to accommodate the power calculation (Kong et al., 2003; Manatunga et al., 2001).

An alternative approach for calculating the power and sample size is to conduct a simulation study. Simulation-based power analysis tends to be more flexible in its accommodation of the trial parameters, including ICC values, number of clusters, cluster sizes, and desired significance levels, etc (Turner et al., 2005, 2004). In the current research context, it is possible to take into account of the influences of historical data, in ways described in the previous chapter. Similar approaches have been used by others, eg. Wang and Gelfand (2002) in a fully Bayesian context. They proposed to elicit an informative prior on the treatment effect from the historical information, which was referred to as the sampling prior, and then developed a simulation-based algorithm to determine the sample size under the pre-specified sampling prior. To implement, the treatment effects are generated from the sampling prior, then data replicates are generated from the model under a given sample size. Then they fitted the same model under different data replicates, the new priors were referred to as the fitting prior. Analytical power was estimated from the number of times of rejecting the null hypothesis under the simulated data.

Notably, this fully Bayesian approach does not preserve the interpretation of the traditional trial concepts, such as type I and type II error rates because the treatment itself is viewed as random. Along the same vein, other Bayesian inference approaches have been developed under the decision theoretical framework (Pezeshk, 2003), which built upon the concepts of the average coverage criterion (ACC), average length criterion (ALC), Bayes factor, loss and utility functions, and etc (Berger, 2013; Weiss, 1997).

To retain the basic frequentist trial concepts, Stamey et al. (2013) proposed to use a point mass prior for the treatment effect as the sampling prior. In its essence,

this modified method is not fully Bayesian. Instead, it only used Bayesian techniques to design a new trial under the traditional criteria.

In this dissertation, we propose a simulation scheme by which we could estimate the analytical power, while incorporating the previous trial data through a power prior, along the lines of Hobbs and Carlin (2007) and Ibrahim and Chen (2000).

3.2 Designing Trials Using Bayesian Power Priors

Power prior has been used previously in designing various types of clinical trials, including non-inferiority trials, one-sided superiority trials, and even sequential trials (Chen et al., 2014, 2011, 2014; Ibrahim et al., 2012). To the best of our knowledge, the method has not been used to design superiority cluster randomized trials. Assuming that one historical dataset $\boldsymbol{D}_0$ is available, following Model (2.5), we aim at developing a simulation scheme through which we could estimate the analytical power for a two-arm cluster randomized trial with a binary outcome.

In this context, the sample size consists of cluster number and cluster size. For the purpose of illustration, we determine the number of clusters with fixed cluster size. We assume the current data $\boldsymbol{D}$ come from Model (2.4), which involves three parameters, β_0, β_1 and τ_b, where β_1 is the treatment effect and the other two parameters can be considered as nuisance ones. Then the two-sided superiority test is

$$\mathrm{H}_0 : \mathrm{E}(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0) = 0 \quad \text{v.s.} \quad \mathrm{H}_1 : \mathrm{E}(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0) \neq 0. \tag{3.1}$$

We determine the minimal cluster number that controls the type I error rate for less than or equal to α_1, and power greater than or equal to α_2. Here α_1 and α_2 are user-specified threshold values, and typical values are 0.05 and 0.8, respectively.

We assume the sampling prior $\pi^s(\beta_0, \beta_1, \tau_b) = \pi^s(\beta_0)\pi^s(\beta_1)\pi^s(\tau_b)$, where $\pi^s(\beta_1)$ is a point mass prior $\pi_0^s(\beta_1) = 1$ under the null hypothesis for calculation of type I error rate, and any prior $\pi_1^s(\beta_1)$ with its support as a subset of $\{\beta_1 : \beta_1 \neq 0\}$ in the calculation of power. A simple example is the point mass prior $\pi_1^s(\beta_1) = 1$ when $\beta_1 = \beta_1^\star$, where $\beta_1^\star$ is any nonzero value. No constraints are imposed on $\pi^s(\beta_0)$ and $\pi^s(\tau_b)$, and they are typically specified as point mass priors. In addition, we assume the fitting prior is $\pi^f(\beta_0, \beta_1, \tau_b) = \pi^f(\beta_0)\pi^f(\beta_1)\pi^f(\tau_b)$, where $\pi^f(\beta_1)$ is the power prior (2.11), and no constraints are imposed on $\pi^f(\beta_0)$ and $\pi^f(\tau_b)$, and they are usually specified as non-informative priors.

With the sampling prior $\pi^s(\beta_0, \beta_1, \tau_b)$, the current data $\boldsymbol{D}$ can be generated from Model (2.4), then the power prior analysis can be conducted to the data $\boldsymbol{D}$ with the fitting prior $\pi^f(\beta_0, \beta_1, \tau_b)$, and therefore the posterior of β_1 (2.12) is yielded. From this posterior, the $(1-\alpha) \times 100\%$ HPD interval of β_1 can be determined, and it is used to complete the inference (3.1). We reject H_0 if the HPD interval does not cover 0, and retain it otherwise.

This rejection probability is taken as the design quantity, which can be written as

$$\beta_{1s}^{I} = \mathrm{E}_s\{\mathrm{IND}(P(\beta_1 > 0|\boldsymbol{D}, \pi^f) > 1 - \frac{\alpha}{2}) + \mathrm{IND}(P(\beta_1 < 0|\boldsymbol{D}, \pi^f) > 1 - \frac{\alpha}{2})\}, \quad (3.2)$$

where I is the experimental cluster number, IND is the indicator function, $P(\beta_1 > 0|\boldsymbol{D}, \pi^f) > 1 - \frac{\alpha}{2}$ and $P(\beta_1 < 0|\boldsymbol{D}, \pi^f) > 1 - \frac{\alpha}{2}$ are respectively the event that the $\frac{\alpha}{2}$th quantile of β_1 is greater than 0 and the event that the $(1 - \frac{\alpha}{2})$th quantile of β_1 is smaller than 0. These two events are determined with the posterior (2.12). The expectation $\mathrm{E}_s\{\cdot\}$ is taken with respect to the marginal distribution of $\boldsymbol{D}$ under the sampling prior $\pi^s(\beta_0, \beta_1, \tau_b)$. This design quantity is exactly the type I error rate when $\pi_0^s(\beta_1)$ is used, or the power when $\pi_1^s(\beta_1)$ is used.

For given $0 < \alpha_1 < 1$ and $0 < \alpha_2 < 1$, we compute $I_{\alpha_1} = \min\{I : {\beta_1}_0^I \leq \alpha_1\}$ and $I_{\alpha_2} = \min\{I : {\beta_1}_1^I \geq \alpha_2\}$, where ${\beta_1}_0^I$ and ${\beta_1}_1^I$ are the design quantity (3.2) calculated under the sampling prior $\pi_0^s(\beta_1)$ and $\pi_1^s(\beta_1)$. Therefore, the Bayesian sample size is

$$I_B = \max\{I_{\alpha_1}, I_{\alpha_2}\}. \tag{3.3}$$

However, the design quantity (3.2) does not have a closed form, so no closed form formula for sample size exists in this context. To solve, we develop a simulation based algorithm to search for the Bayesian sample size I_B, as described below.

1. Specify α, α_1, α_2, M and K.
2. Set a value for I, the cluster number.
3. Generate data $\boldsymbol{D}$ from model (2.4), with the sampling prior $\pi^s(\beta_0)\pi^s(\beta_1)\pi^s(\tau_b)$. Use $\pi_0^s(\beta_1)$ to calculate type I error rate ${\beta_1}_0^I$ and use $\pi_1^s(\beta_1)$ to calculate power ${\beta_1}_1^I$.
4. Determine the symmetric and asymmetric discounting fractions (2.8) and (2.9), with $\boldsymbol{D}$ and $\boldsymbol{D}_0$.
5. Construct the power prior of β_1 (2.11) under the historical data $\boldsymbol{D}_0$, for each discounting fraction.

6. Conduct the power prior analysis (2.12) to the current data $\boldsymbol{D}$ with power prior (2.11).
7. Let $\{\beta_1^{(1)}, \ldots, \beta_1^{(M)}\}$ be the posterior samples of β_1 from (2.12). From the posterior samples, calculate

$$\hat{P}_1 = \frac{1}{M}\sum_{m=1}^{M} \text{IND}(\beta_1^{(m)} > 0),$$

and

$$\hat{P}_2 = \frac{1}{M}\sum_{m=1}^{M} \text{IND}(\beta_1^{(m)} < 0),$$

where $\text{IND}(\cdot)$ is the indicator function.
8. Check whether $\hat{P}_1 > 1 - \frac{\alpha}{2}$ or $\hat{P}_2 > 1 - \frac{\alpha}{2}$, and let $R = 1$ if it holds and 0 otherwise.
9. Repeat the steps $3 - 8$ K times, and compute the average of R in these K runs as the rejection proportion, which is an estimate of the design quantity ${\beta_1}_s^I$ in (3.2).
10. If the criteria ${\beta_1}_0^I \leq \alpha_1$ and ${\beta_1}_1^I \geq \alpha_2$ are not satisfied, or I is not the minimum cluster number that makes the criteria hold, try another cluster number and rerun the steps $2 - 9$ until the Bayesian sample size I_B is found.

This Bayesian design is easily extendable to the determination of cluster size with fixed cluster number as well as the case where multiple historical data are available.

3.3 Simulation Studies

A simulation study was designed to illustrate the process of power calculation. We considered a situation where one historical data was available, and we assumed both

current and historical trials were balanced in treatment group size. We considered the determination of (1) the number of clusters with fixed cluster size; and (2) the cluster size with a fixed number of clusters. In both scenarios, the historical trial had $N = 1000, \beta_0' = 0.1, \beta_1' = \log(4)$. The current trial configurations were given in Tables 3.1 and 3.2.

To design a clinical trial (i.e., the current trial), we assumed that the historical data were given. In the simulation study, we first generated the historical data $\boldsymbol{D}_0$ from model (2.5) under a specific parameter setting. Separately we simulated the current trial data. Then we determine the power under specific sample size using the power prior analysis. We did not calculate the Bayesian sample size I_B under each current trial setting due to computation limit. Instead, we implemented the algorithm (without the last step) with some cluster numbers I under each current trial setting in the first scenario and some cluster sizes J under each current trial setting in the second scenario. As a result, the rejection proportions for the null hypothesis $\mathrm{H}_0 : \mathrm{E}(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0) = 0$ could be calculated. We tried cluster number $I = 10, 20, 30, 40$ for each current trial setting in the first scenario, and cluster size $J = 10, 20, 30, 40$ for each current trial setting in the second scenario. Therefore, for each specific current trial setting, we could plot the profile of rejection proportion against cluster number or cluster size, which could be used to propose a new I or J to try further with the simulation based algorithm until I_B is found.

In Tables 3.1 and 3.2, when the current trial setting in the first case was used, the rejection proportion was an estimate of type I error rate, and power otherwise. In the simulation based algorithm, we applied four different discounting fractions, i.e. $0, 1, a^{\mathrm{sym}}$ and a^{asym}. The first two discounting fractions were used for reference. $\pi^f(\beta_0)$

and $\pi^f(\tau_b)$ were non-informative priors, $\alpha = 0.05$, $\alpha_1 = 0.05$, $\alpha_2 = 0.8$, $M = 2,500$ and $K = 1,000$. For the purpose of illustration, the profile of the rejection proportion against cluster number I in the first scenario and that against cluster size J in the second scenario were plotted in Figures $3.1 - 3.2$. From the two Figures, we could see the type I error rate was controlled after the historical information was borrowed with our power prior method, and at the same time, fewer samples were needed to achieve a pre-specified power.

3.4 Example

We assumed that historical trial data in the Example section of last chapter followed model (2.5), and we used it to design a two-arm (balanced) cluster randomized trial with a binary outcome. We conducted the trial design as what we did in the Simulation Studies section above, and the only difference was that we replaced the simulated historical data $\boldsymbol{D}_0$ with this real one. For illustrative purpose, the profile of the rejection proportion against cluster number I in the first scenario and that against cluster size J in the second scenario were plotted in Figures $3.3 - 3.4$. From these two figures, we could read the same message as those from the simulation results.

3.5 Discussion

In this chapter, we used our novel power prior in designing a two-arm cluster randomized trial with binary outcome, and both simulation results and real data analytical results were promising. They both demonstrated that fewer samples were needed to achieve a pre-specified power after the historical information was borrowed with our power prior method, and at the same time, the type I error rate was controlled.

This method, however, had no explicit formula for sample size, and therefore we developed a simulation-based algorithm to determine the sample size. This algorithm was easy to understand in concept, and it was straightforward to operate as well. But it was quite time consuming as it was based on simulations and therefore extensive computations were needed. This drawback would prevent this method from being used in applications. From the details of this algorithm described in this chapter, we could see it consisted of several similar computations, and then the results from them were pooled to yield the final results. So we used the parallel computing approach to solve this problem. In the parallel computing program, the similar computation tasks could be done simultaneously on different cores of a computer, and therefore the computation time was decreased compared to serial computing approach. The decrease in computing time was inversely proportional to the number of cores available on the computer.

This power prior method has demonstrated its use in the design and analysis of two-arm cluster randomized trial with binary outcome. Its value, however, would be too limited if we could not extend it to more general situations, such as multi-arm cluster randomized trials with outcome following arbitrary exponential family distributions. This aim would be covered in the next chapter.

Table 3.1: Current trial setting for the determination of cluster number in two-arm trials with binary outcome

Case	J	β_0	β_1	τ_b
1	20	Unif$(-0.1, 0.1)$	$\log(1)$	Unif$(0.463, 1)$
2	20	Unif$(-0.1, 0.1)$	$\log(2)$	Unif$(0.463, 1)$
3	20	Unif$(-0.1, 0.1)$	$\log(4)$	Unif$(0.463, 1)$
4	20	Unif$(-0.1, 0.1)$	Unif$(\log(2), \log(4))$	Unif$(0.463, 1)$

Table 3.2: Current trial setting for the determination of cluster size in two-arm trials with binary outcome

Case	I	β_0	β_1	τ_b
1	20	Unif$(-0.1, 0.1)$	$\log(1)$	Unif$(0.463, 1)$
2	20	Unif$(-0.1, 0.1)$	$\log(2)$	Unif$(0.463, 1)$
3	20	Unif$(-0.1, 0.1)$	$\log(4)$	Unif$(0.463, 1)$
4	20	Unif$(-0.1, 0.1)$	Unif$(\log(2), \log(4))$	Unif$(0.463, 1)$

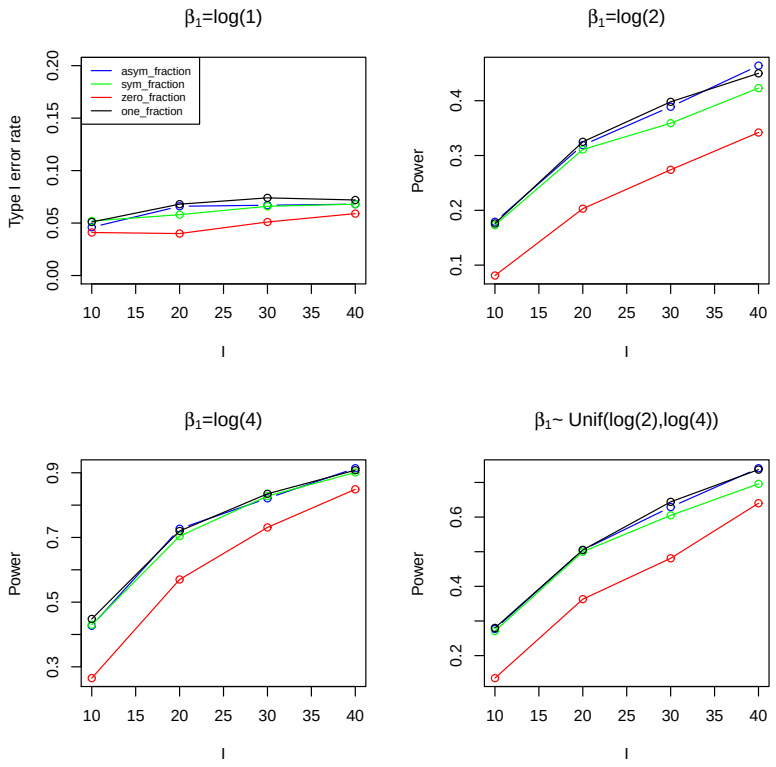

Figure 3.1: In two-arm trials with binary outcome, the profile of type I error rate and power against cluster number I where the cluster size $J = 20$, $\beta_0 \sim \text{Unif}(-0.1, 0.1)$, $\tau_b \sim \text{Unif}(0.463, 1)$ in current trial and the arm size $N = 1,000$, $\beta_0' = 0.1$, $\beta_1' = \log(4)$ in historical trial

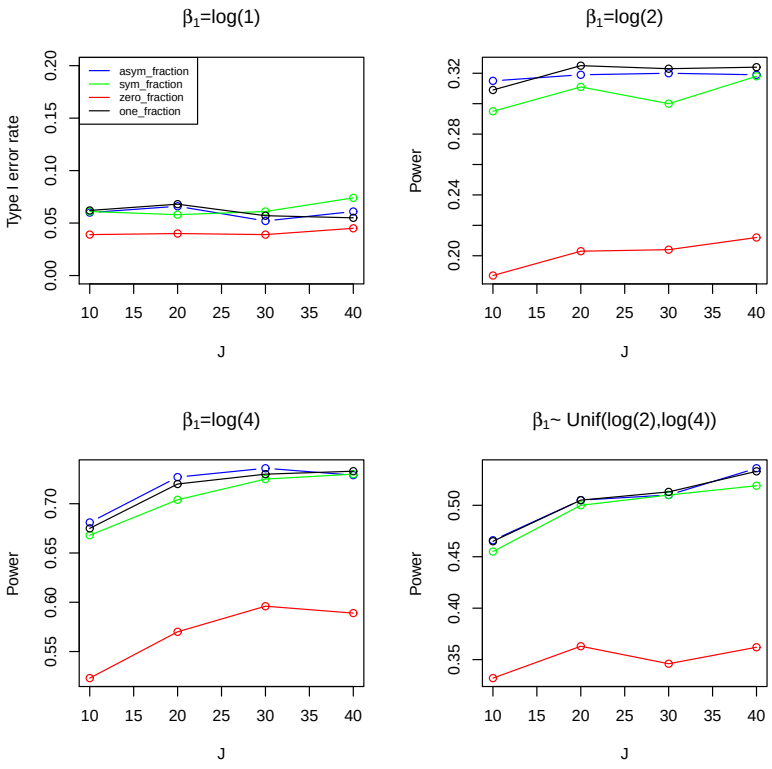

Figure 3.2: In two-arm trials with binary outcome, the profile of type I error rate and power against cluster size J where the cluster number $I = 20$, $\beta_0 \sim \text{Unif}(-0.1, 0.1)$, $\tau_b \sim \text{Unif}(0.463, 1)$ in current trial and the arm size $N = 1,000$, $\beta_0' = 0.1$, $\beta_1' = \log(4)$ in historical trial

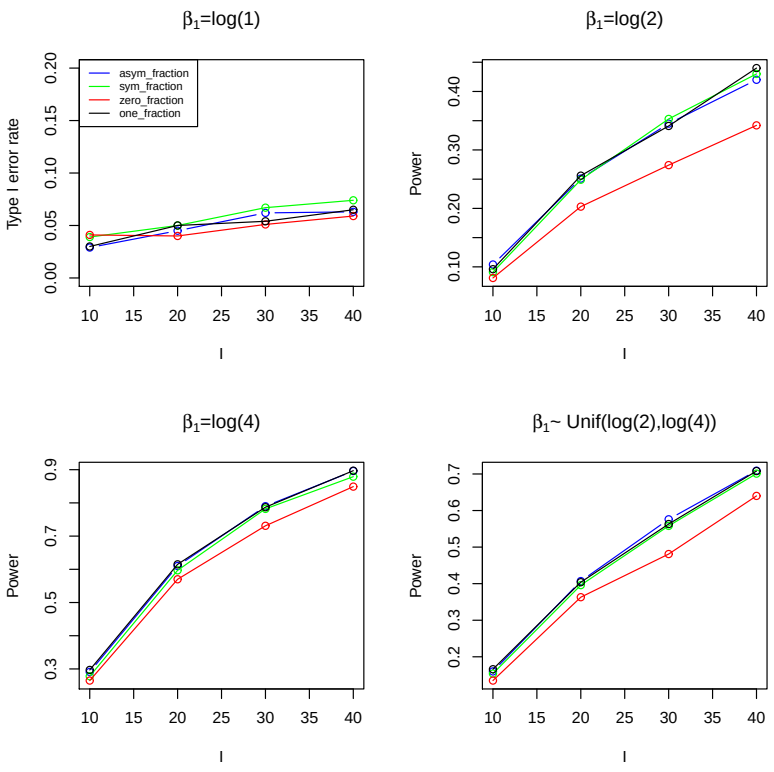

Figure 3.3: In two-arm trials with binary outcome, the profile of type I error rate and power against cluster number I where the cluster size $J = 20$, $\beta_0 \sim \text{Unif}(-0.1, 0.1)$, $\tau_b \sim \text{Unif}(0.463, 1)$ in current trial and a real historical trial is used

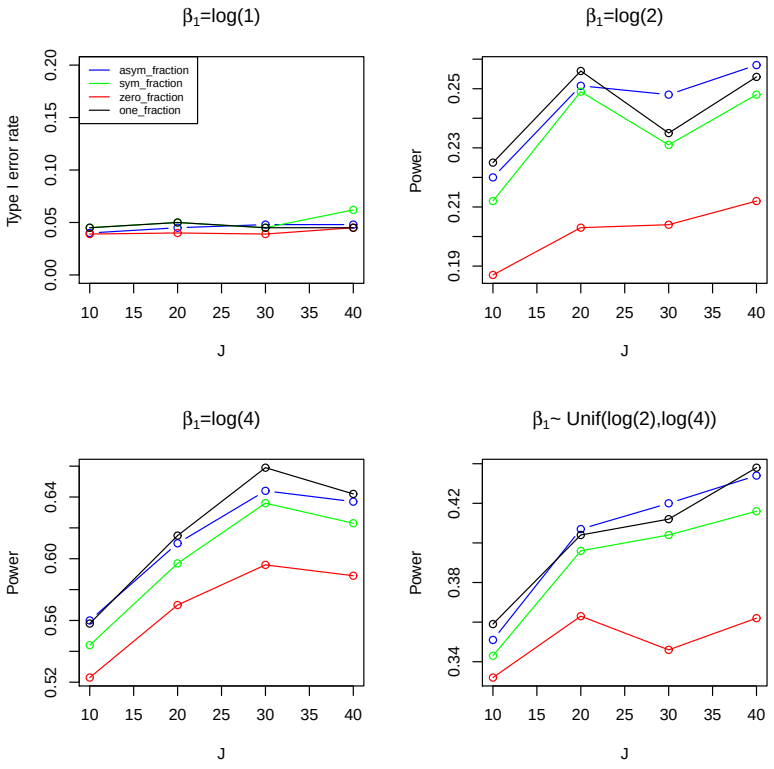

Figure 3.4: In two-arm trials with binary outcome, the profile of type I error rate and power against cluster size J where the cluster number $I = 20$, $\beta_0 \sim \text{Unif}(-0.1, 0.1)$, $\tau_b \sim \text{Unif}(0.463, 1)$ in current trial and a real historical trial is used

Chapter 4

Further Extensions

In this chapter, we extend the proposed method to situations of multi-arm cluster randomized trials and outcomes following the exponential family distributions, and thus including continuous, count, binary, and other categorical outcomes. We also extend the method to accommodate both balanced and imbalanced designs. Modeling details and computational techniques are largely parallel to those presented in the previous chapters.

4.1 Cluster Randomized Trials: Notation

Suppose that the current trial consists of one control arm and $P-1$ different treatment arms, and the numbers of clusters in the control arm and the treatment arms are I_0 and $I_p, p = 1, 2, \ldots, P-1$. So the total number of clusters is $I = I_0 + I_1 + \ldots + I_{P-1}$. Let J_i be the number of subjects in the ith cluster, and Y_{ij} the outcome from the jth subjects in the ith cluster following an exponential family distribution. We additionally use $P-1$ indicator variables $X_{1i}, X_{2i}, \ldots, X_{P-1i}$ to indicate the treatment assignment for the ith cluster. Specifically, if the ith cluster is in the pth treatment arm, then all indicator variables take value 0 except for $X_{pi} = 1$. Under such notation, the current trial data can be written as $\boldsymbol{D} = \{(Y_{ij}, X_{1i}, X_{2i}, \ldots, X_{P-1i}) : i = 1, 2, \ldots, I, j = 1, 2, \ldots, J_i\}$.

For the historical trial, we assume that it has one control group and $P-1$ similar treatment groups. Let l be the treatment group indicator, with 0 for the control group and $1, 2, \ldots, P-1$ for the corresponding treatment groups. Suppose that the historical data come in a summary fashion. We let N_l be the number of subjects in the lth arm, Z_l be the summary statistic of individual outcomes in the lth treatment group. In particular, Z_l is a sufficient statistic of the mean of the lth group. As in the current trial, we use indicator variables $X'_{1l}, X'_{2l}, \ldots, X'_{P-1l}$ to indicates the treatment assignment. For the lth group, we all indicator variables take value 0 except for $X'_{pl} = 1$. So the historical data can be written as $\boldsymbol{D}_0 = \{(Z_l, N_l, X'_{1l}, X'_{2l}, \ldots, X'_{P-1l}) : l = 0, 1, \ldots, P-1\}$.

Suppose that the trial outcomes follow an exponential family distribution

$$f(Y_{ij}|b_i) = \exp\{\frac{Y_{ij}\eta_i - d(\eta_i)}{a(\phi)} + c(Y_{ij}, \phi)\}, \tag{4.1}$$

where η_i is the natural parameter, and ϕ a nuisance parameter. Let μ_i be the conditional mean of Y_{ij} given b_i, and it is related to η_i through a monotone invertible link function $k(\cdot)$.

The outcomes are linked to the independent variables through the following model

$$\eta_i = k(\mu_i) = \beta_0 + \boldsymbol{X}_i^T\boldsymbol{\beta} + b_i, \tag{4.2}$$

where $\boldsymbol{X}_i^T = (X_{1i}, X_{2i}, \ldots, X_{P-1i})$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_{P-1})^T$ is the $(P-1)\times 1$ coefficient vector, and b_i is the random effect associated with the ith cluster. For narrative simplicity, we write $\boldsymbol{b} = (b_1, b_2, \ldots, b_I)$. As in the standard mixed effect generalized linear models, we assume $b_i \sim \mathrm{N}(0, \tau_b)$, $\beta_0 \sim \mathrm{N}(\mu_{\beta_0}, \tau_{\beta_0})$, $\beta_p \sim N(\mu_{\beta_p}, \tau_{\beta_p})$ and $\tau_b \sim$

Gamma(κ_b, ν_b). The nuisance parameter ϕ is not the parameter of primary interest and its functional form varies among different members of the exponential family. For example, ϕ is the precision parameter in normal distribution, and shape parameter in Beta distribution. Without loss of generality, we assume that the prior for ϕ is $f(\phi|\boldsymbol{\zeta})$, where $\boldsymbol{\zeta}$ is the hyper-parameter vector. Here β_p represents the treatment effect of the pth treatment arm under $\boldsymbol{D}$.

Similarly, for the historical data $\boldsymbol{D}_0$, we assume that the individual outcomes follow the same distribution as does the current trial outcome. We write the individual outcomes in the lth group as Z_{il}, $i = 1, 2, \ldots, N_l$ and $l = 0, 1, \ldots, P-1$

$$g(Z_{il}) = \exp\{\frac{Z_{il}\eta_l' - d(\eta_l')}{a(\phi')} + c(Z_{il}, \phi')\}, \tag{4.3}$$

where η_l' is the natural parameter, and ϕ' is the nuisance parameter. Therefore, the joint distribution of $(Z_{1l}, Z_{2l}, \ldots, Z_{N_l l})$ is

$$\begin{aligned} g(Z_{1l}, Z_{2l}, \ldots, Z_{N_l l}) &= \exp\{\frac{\eta_l' \sum_{i=1}^{N_l} Z_{il} - N_l d(\eta_l')}{a(\phi')} + \sum_{i=1}^{N_l} c(Z_{il}, \phi')\} \\ &= \exp\{\frac{Z_l\eta_l' - N_l d(\eta_l')}{a(\phi')}\}\exp\{\sum_{i=1}^{N_l} c(Z_{il}, \phi')\}. \end{aligned} \tag{4.4}$$

The sample sum Z_l is a sufficient statistic for the natural parameter η_l' by the factorization theorem. We then derive the probability density function of Z_l. From Equation (4.4), we have

$$g(Z_l) = \int \cdots \int \exp\{\frac{Z_l\eta_l' - N_l d(\eta_l')}{a(\phi')}\}\exp\{\sum_{i=1}^{N_l-1} c(Z_{il},\phi') + c(Z_l - \sum_{i=1}^{N_l-1} Z_{il},\phi')\}\, dZ_{1l}\ldots\, dZ_{N_l-1l}, \tag{4.5}$$

under the transformation $Z_{1l} = Z_{1l}, \ldots, Z_{N_l-1l} = Z_{N_l-1l}, Z_{N_l l} = Z_l - \sum_{i=1}^{N_l-1} Z_{il}$ as the Jacobian determinant is 1.

For normal, count and binary outcomes, the distributions of the sufficient statistic Z_l are $\mathrm{N}(N_l\eta_l', N_l a(\phi'))$, $\mathrm{Poisson}(N_l\exp(\eta_l'))$ and $\mathrm{Binomial}(N_l, \mathrm{logit}^{-1}(\eta_l'))$, respectively. Let μ_l' be the mean of Z_{il}, or equivalently the mean of $\frac{Z_l}{N_l}$, and it is related to η_l' through the monotone invertible link function $k(\cdot)$.

Consider the following model

$$\eta_l' = k(\mu_l') = \beta_0' + \boldsymbol{X}_l'^T\boldsymbol{\beta'}, \tag{4.6}$$

where $\boldsymbol{X}_l'^T = (X_{1l}', X_{2l}', \ldots, X_{P-1l}')$, $\boldsymbol{\beta'} = (\beta_1', \beta_2', \ldots, \beta_{P-1}')^T$ is the $(P-1)\times 1$ coefficient vector. We further assume $\beta_0' \sim \mathrm{N}(\mu_{\beta_0'}, \tau_{\beta_0'})$, $\beta_p' \sim \mathrm{N}(\mu_{\beta_p'}, \tau_{\beta_p'})$, and the prior for the nuisance parameter ϕ' is $g(\phi'|\boldsymbol{\zeta'})$, where $\boldsymbol{\zeta'}$ is the hyper-parameter vector, and β_p' stands for the treatment effect of the pth treatment group under $\boldsymbol{D}_0$.

4.2 Determination of Discounting Fractions

All priors for parameters in models (4.2) and (4.6) are assumed to be non-informative. In separate data analysis, the posteriors of $\boldsymbol{\beta}$ and $\boldsymbol{\beta'}$, denoted by $f(\boldsymbol{\beta}|\boldsymbol{D})$ and $g(\boldsymbol{\beta'}|\boldsymbol{D}_0)$,

are evaluated, in the presence of closed forms. To implement, we use the MH algorithm within Gibbs sampling to ascertain the corresponding posterior samples.

Since the two posteriors are derived under non-informative priors, they are not influenced by external data other than the current and historical information. From this, the symmetric and asymmetric KL divergence measures are defined to quantify the similarity of the two data sources

$$D_{KL}^{\text{sym}}(f||g) = \text{E}_f\{\log(\frac{f}{g})\} + \text{E}_g\{\log(\frac{g}{f})\}, \tag{4.7}$$

$$D_{KL}^{\text{asym}}(f||g) = \text{E}_f\{\log(\frac{f}{g})\}, \tag{4.8}$$

where f and g stand for $f(\boldsymbol{\beta}|\boldsymbol{D})$ and $g(\boldsymbol{\beta'}|\boldsymbol{D}_0)$, and $\text{E}_f\{\cdot\}$ and $\text{E}_g\{\cdot\}$ are the expectations taken with respect to f and g, respectively.

Then we propose the following discounting fractions

$$a^{\text{sym}} = e^{-D_{KL}^{\text{sym}}(f||g)} = e^{-D_{KL}^{\text{sym}}(g||f)}. \tag{4.9}$$

$$a^{\text{asym}} = e^{-D_{KL}^{\text{asym}}(f||g)}, \tag{4.10}$$

The two KL divergence measures are not easy to compute by definition, we therefore use the k-Nearest Neighbor (k-NN) algorithm for calculation.

4.3 Estimation of Treatment Effect

We assume the pth treatment in the current trial is only the same or similar to that in the historical trial, and therefore, for the estimation of β_p, it is reasonable to allow

the historical information be borrowed only through β_p'. Under this assumption, we assume β_p and β_p' are exchangeable. Denote $\boldsymbol{\beta}_p = (\beta_p, \beta_p')$, and we assume it follows a bivariate normal distribution $h_p(\boldsymbol{\beta}_p|\boldsymbol{m}_p, \boldsymbol{\Lambda}_p)$, with a mean vector $\boldsymbol{m}_p = \mu_p(1,1)'$ and a 2×2 precision matrix $\boldsymbol{\Lambda}_p$. Here, the precision and correlation parameters are τ_p and ρ_p, respectively.

To complete the model specification, we assume the following hyper-prior distributions for μ_p, τ_p and ρ_p

$$h_p(\mu_p|a_p, R_p) = \sqrt{\frac{R_p}{2\pi}} e^{-\frac{R_p}{2}(\mu_p - a_p)^2},$$

$$h_p(\tau_p|\kappa_p, \nu_p) = \frac{\nu_p^{\kappa_p}}{\Gamma(\kappa_p)} \tau_p^{\kappa_p - 1} e^{-\nu_p \tau_p},$$

$$h_p(\rho_p|c_p, d_p) = \frac{\rho_p^{c_p - 1}(1-\rho_p)^{d_p - 1}}{B(c_p, d_p)},$$

where $\Gamma(\cdot)$ and $B(\cdot, \cdot)$ are respectively Gamma and Beta functions.

The power prior of $\boldsymbol{\beta}$ can therefore be written as

$$\begin{aligned} h(\boldsymbol{\beta}|\boldsymbol{D}_0, a) \propto \int \cdots \int \{ \int (L(\beta_0', \boldsymbol{\beta}'|\boldsymbol{D}_0)^a g(\beta_0')\, d\beta_0' \} \prod_{p=1}^{P-1} \{ h_p(\boldsymbol{\beta_p}|\boldsymbol{m}_p, \boldsymbol{\Lambda}_p) \\ h_p(\mu_p|a_p, R_p) h_p(\tau_p|\kappa_p, \nu_p) h_p(\rho_p|c_p, d_p) \}\, d\boldsymbol{\beta}'\, d\boldsymbol{\mu}\, d\boldsymbol{\tau}\, d\boldsymbol{\rho}, \end{aligned} \tag{4.11}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_{P-1})$, $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_{P-1})$ and $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_{P-1})$, $L(\beta_0', \boldsymbol{\beta}'|\boldsymbol{D}_0)$ and $g(\beta_0')$ are the likelihood and prior density function of β_0' and $\boldsymbol{\beta}'$ in Model (4.6). From the power prior (4.11), we derive the posterior function of $\boldsymbol{\beta}$ as

follows

$$h(\boldsymbol{\beta}|\boldsymbol{D},\boldsymbol{D}_0,a) \propto \int\cdots\int \{L(\beta_0,\boldsymbol{\beta}|\boldsymbol{D},\boldsymbol{b})f(\boldsymbol{b}|\tau_b)f(\beta_0)f(\tau_b)\,d\boldsymbol{b}\,d\beta_0\,d\tau_b\}h(\boldsymbol{\beta}|\boldsymbol{D}_0,a), \tag{4.12}$$

where $L(\beta_0,\boldsymbol{\beta}|\boldsymbol{D},\boldsymbol{b})$, $f(\boldsymbol{b}|\tau_b)$, $f(\beta_0)$ and $f(\tau_b)$ are the likelihood and prior density functions in Model (4.2).

With the posterior (4.12), we can answer various questions about the treatment effects. For examples,

1. Does the pth treatment has a nonzero effect? The hypothesis is equivalent to testing $\mathrm{H}_0 : E(\beta_p|\boldsymbol{D},\boldsymbol{D}_0) = 0$, for a given $p = 1, 2, \ldots, P-1$.
2. Is the effect of the pth treatment significantly different from that of the kth treatment? The hypothesis is equivalent to testing $\mathrm{H}_0 : E(\gamma_{pk}|\boldsymbol{D},\boldsymbol{D}_0) = 0$, where $\gamma_{pk} = \beta_p - \beta_k$, for $k = 1, 2, \ldots, P-1$, where $k \neq p$.
3. Do the $P-1$ treatments have a similar effect? The hypothesis is equivalent to testing $\mathrm{H}_0 : E(\beta_1|\boldsymbol{D},\boldsymbol{D}_0) = E(\beta_2|\boldsymbol{D},\boldsymbol{D}_0) = \cdots = E(\beta_{P-1}|\boldsymbol{D},\boldsymbol{D}_0)$.

In case of all pairwise comparisons, one could adjust for inflated type I error rates associated with multiple test, following an established method (Hsu, 1996).

For the first hypothesis, we rely on the posterior of β_p

$$h(\beta_p|\boldsymbol{D},\boldsymbol{D}_0,a) = \int\cdots\int h(\boldsymbol{\beta}|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\boldsymbol{\beta}_{-p}, \tag{4.13}$$

where $\boldsymbol{\beta}_{-p}$ is the vector $\boldsymbol{\beta}$ with the pth element β_p removed. We use $h(\beta_p|\boldsymbol{D},\boldsymbol{D}_0,a)$ to estimate β_p and complete the inference.

For the second hypothesis, we rely on the posterior of γ_{pk}, which we derive from the joint posterior of (β_p, β_k)

$$h(\beta_p, \beta_k | \boldsymbol{D}, \boldsymbol{D}_0, a) = \int \cdots \int h(\boldsymbol{\beta} | \boldsymbol{D}, \boldsymbol{D}_0, a)\, d\boldsymbol{\beta}_{-pk}, \tag{4.14}$$

where $\boldsymbol{\beta}_{-pk}$ is the vector $\boldsymbol{\beta}$ with the pth and kth elements β_p and β_k removed. The posterior of γ_{pk} can be obtained as

$$h(\gamma_{pk} | \boldsymbol{D}, \boldsymbol{D}_0, a) = \int h(\beta_p, \beta_p - \gamma_{pk} | \boldsymbol{D}, \boldsymbol{D}_0, a)\, d\beta_p \tag{4.15}$$

under the transformation $\beta_p = \beta_p$, $\gamma_{pk} = \beta_p - \beta_k$. Then we can use $h(\gamma_{pk} | \boldsymbol{D}, \boldsymbol{D}_0, a)$ to estimate γ_{pk} and complete the inference.

For the third hypothesis, we first derive the posterior distributions for all β_p, i.e. $h(\beta_p | \boldsymbol{D}, \boldsymbol{D}_0, a), p = 1, 2, \ldots, P-1$, and then calculate the $(1 - \frac{\alpha}{P-1}) \times 100\%$ HPD intervals of β_p and use them to complete the inference. Specifically, the null hypothesis is retained only when all HPD intervals cover 0.

For a given scientific question concerning the treatment effects, we should express it as a contrast containing $\boldsymbol{\beta}$. We will then derive the posterior of that contrast based on $h(\boldsymbol{\beta} | \boldsymbol{D}, \boldsymbol{D}_0, a)$ under an appropriate variable transformation. We will then use that posterior to achieve estimation and inference. This said, none of the aforementioned posteriors have closed forms, and so we will use an MH algorithm within Gibbs sampling to draw posterior samples from them for calculation of the summary statistics of the posterior samples.

4.4 Algorithms

4.4.1 The K-Nearest Neighbor Algorithm

We used the k-NN algorithm to calculate the symmetric and asymmetric KL divergence (4.7) and (4.8) with the posterior samples of $\boldsymbol{\beta}$ from $f(\boldsymbol{\beta}|\boldsymbol{D})$ and those of $\boldsymbol{\beta'}$ from $g(\boldsymbol{\beta'}|\boldsymbol{D}_0)$. We write the posterior samples of $\boldsymbol{\beta}$ and $\boldsymbol{\beta'}$ as $\{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(n)}\}$ and $\{\boldsymbol{\beta'}^{(1)}, \boldsymbol{\beta'}^{(2)}, \ldots, \boldsymbol{\beta'}^{(m)}\}$, respectively.

Let $\rho_k(i)$ be the Euclidean distance between $\boldsymbol{\beta}^{(i)}$ and its k-NN in $\{\boldsymbol{\beta}^{(j)}\}_{j\neq i}$. Specifically, the k-NN of x in $\{z_1, z_2, \ldots, z_n\}$ is $z_{i(k)}$ where $i(1), i(2), \ldots, i(n)$ are such that

$$||x - z_{i(1)}|| \leq ||x - z_{i(2)}|| \leq \ldots \leq ||x - z_{i(n)}||.$$

The distance from $\boldsymbol{\beta}^{(i)}$ to its k-NN in $\{\boldsymbol{\beta'}^{(1)}, \boldsymbol{\beta'}^{(2)}, \ldots, \boldsymbol{\beta'}^{(m)}\}$ is denoted by $\nu_k(i)$, then the asymmetric KL divergence $D_{\mathrm{KL}}^{\mathrm{asym}}(f||g)$ can be calculated as

$$\hat{D}_{\mathrm{KL}}^{\mathrm{asym}}(f||g) = \frac{1}{n}\sum_{i=1}^{n} \log(\frac{\nu_k(i)}{\rho_k(i)}) + \log(\frac{m}{n-1}). \tag{4.16}$$

Similarly, we let $\kappa_k(j)$ be the Euclidean distance between $\boldsymbol{\beta'}^{(j)}$ and its k-NN in $\{\boldsymbol{\beta'}^{(i)}\}_{i\neq j}$, and the Euclidean distance from $\boldsymbol{\beta'}^{(j)}$ to its k-NN in $\{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(n)}\}$ be $\lambda_k(j)$, then the symmetric KL divergence $D_{\mathrm{KL}}^{\mathrm{sym}}(f||g)$ can be calculated as below,

$$\hat{D}_{\mathrm{KL}}^{\mathrm{sym}}(f||g) = \frac{1}{n}\sum_{i=1}^{n} \log(\frac{\nu_k(i)}{\rho_k(i)}) + \log(\frac{m}{n-1}) + \frac{1}{m}\sum_{j=1}^{m} \log(\frac{\lambda_k(j)}{\kappa_k(j)}) + \log(\frac{n}{m-1}). \tag{4.17}$$

As a result, we have the symmetric and asymmetric discounting fractions as

$$\hat{a}^{\mathrm{sym}} = e^{-\hat{D}_{\mathrm{KL}}^{\mathrm{sym}}(f||g)} = e^{-\hat{D}_{\mathrm{KL}}^{\mathrm{sym}}(g||f)}.$$

$$\hat{a}^{\text{asym}} = e^{-\hat{D}_{\text{KL}}^{\text{asym}}(f||g)},$$

4.4.2 The Metropolis Hastings within Gibbs Algorithm

The Gibbs Component of the Hybrid Algorithm

To draw posterior samples of $\boldsymbol{\beta}$ from the posterior $f(\boldsymbol{\beta}|\boldsymbol{D})$, we fit the current data $\boldsymbol{D}$ with Model (4.2). The joint posterior of $\boldsymbol{\beta}$, β_0, $\boldsymbol{b}$, τ_b and ϕ is,

$$\begin{aligned}
&f(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi|\boldsymbol{D}) = \frac{f(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{D})}{f(\boldsymbol{D})} \\
&= \frac{f(\boldsymbol{Y}|\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \phi) f(\boldsymbol{\beta}) f(\beta_0) f(\boldsymbol{b}|\tau_b) f(\tau_b) f(\phi|\boldsymbol{\zeta})}{\int \cdots \int f(\boldsymbol{Y}|\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \phi) f(\boldsymbol{\beta}) f(\beta_0) f(\boldsymbol{b}|\tau_b) f(\tau_b) f(\phi|\boldsymbol{\zeta})\, d\boldsymbol{\beta}\, d\beta_0\, d\boldsymbol{b}\, d\tau_b\, d\phi} \\
&\propto f(\boldsymbol{Y}|\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \phi) f(\boldsymbol{\beta}) f(\beta_0) f(\boldsymbol{b}|\tau_b) f(\tau_b) f(\phi|\boldsymbol{\zeta}) \\
&= \left\{ \prod_{i=1}^{I} \prod_{j=1}^{J_i} \exp\{ \frac{Y_{ij}(\beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta} + b_i) - d(\beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta} + b_i)}{a(\phi)} + c(Y_{ij}, \phi) \} \right\} \\
&\left\{ \prod_{p=1}^{P-1} \sqrt{\frac{\tau_{\beta_p}}{2\pi}} e^{-\frac{\tau_{\beta_p}}{2}(\beta_p - \mu_{\beta_p})^2} \right\} \left\{ \sqrt{\frac{\tau_{\beta_0}}{2\pi}} e^{-\frac{\tau_{\beta_0}}{2}(\beta_0 - \mu_{\beta_0})^2} \right\} \left\{ \prod_{i=1}^{I} \sqrt{\frac{\tau_b}{2\pi}} e^{-\frac{1}{2} b_i^2 \tau_b} \right\} \\
&\left\{ \frac{\nu_b^{\kappa_b}}{\Gamma(\kappa_b)} \tau_b^{\kappa_b - 1} e^{-\nu_b \tau_b} \right\} f(\phi|\boldsymbol{\zeta}) \\
&\propto \left\{ \prod_{i=1}^{I} \prod_{j=1}^{J_i} \exp\{ \frac{Y_{ij}(\beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta} + b_i) - d(\beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta} + b_i)}{a(\phi)} + c(Y_{ij}, \phi) \} \right\} \\
&\left\{ \prod_{p=1}^{P-1} e^{-\frac{\tau_{\beta_p}}{2}(\beta_p - \mu_{\beta_p})^2} \right\} \left\{ e^{-\frac{\tau_{\beta_0}}{2}(\beta_0 - \mu_{\beta_0})^2} \right\} \left\{ \prod_{i=1}^{I} \sqrt{\tau_b} e^{-\frac{1}{2} b_i^2 \tau_b} \right\} \\
&\left\{ \tau_b^{\kappa_b - 1} e^{-\nu_b \tau_b} \right\} f(\phi|\boldsymbol{\zeta}),
\end{aligned}$$

where $\boldsymbol{Y} = (Y_{11}, \ldots, Y_{1J_1}, \ldots, Y_{I1}, \ldots, Y_{IJ_I})$. Therefore, the full conditionals of $\boldsymbol{\beta}$,β_0, $\boldsymbol{b}$, τ_b and ϕ are

$$
\begin{aligned}
&f(\boldsymbol{\beta}|\boldsymbol{D},\beta_0,\boldsymbol{b},\tau_b,\phi)\\
&=f(\boldsymbol{\beta}|\boldsymbol{D},\beta_0,\boldsymbol{b},\phi)\\
&=\frac{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}{f(\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}\\
&=\frac{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}{\int\cdots\int f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})\,d\boldsymbol{\beta}}\\
&=\frac{\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{\prod_{p=1}^{P-1}e^{-\frac{\tau_{\beta_p}}{2}(\beta_p-\mu_{\beta_p})^2}\right\}}{\int\cdots\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{\prod_{p=1}^{P-1}e^{-\frac{\tau_{\beta_p}}{2}(\beta_p-\mu_{\beta_p})^2}\right\}d\boldsymbol{\beta}}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{\prod_{p=1}^{P-1}e^{-\frac{\tau_{\beta_p}}{2}(\beta_p-\mu_{\beta_p})^2}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&f(\beta_0|\boldsymbol{D},\boldsymbol{\beta},\boldsymbol{b},\tau_b,\phi)\\
&=f(\beta_0|\boldsymbol{D},\boldsymbol{\beta},\boldsymbol{b},\phi)\\
&=\frac{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}{f(\boldsymbol{\beta},\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}\\
&=\frac{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}{\int f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})\,d\beta_0}\\
&=\frac{\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\}}{\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\}d\beta_0}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&f(\boldsymbol{b}|\boldsymbol{D},\boldsymbol{\beta},\beta_0,\tau_b,\phi)\\
&=\frac{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}{f(\boldsymbol{\beta},\beta_0,\tau_b,\phi|\boldsymbol{D})}\\
&=\frac{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}{\int\cdots\int f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})\,d\boldsymbol{b}}\\
&=\frac{\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}}{\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}d\boldsymbol{b}}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\},
\end{aligned}
$$

$$
\begin{aligned}
& f(\tau_b|\boldsymbol{D},\boldsymbol{\beta},\beta_0,\boldsymbol{b},\phi) \\
&= f(\tau_b|\boldsymbol{b}) \\
&= \frac{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\phi|\boldsymbol{D})} \\
&= \frac{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}{\int f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})\,d\tau_b} \\
&= \frac{\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}}{\int\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}\,d\tau_b} \\
&\propto \left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\} \\
&\sim \text{Gamma}(\kappa_b+\frac{I}{2},\nu_b+\frac{\sum_i^I b_i^2}{2}),
\end{aligned}
$$

$$
\begin{aligned}
& f(\phi|\boldsymbol{D},\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b) \\
&= f(\phi|\boldsymbol{D},\boldsymbol{\beta},\beta_0,\boldsymbol{b}) \\
&= \frac{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b|\boldsymbol{D})} \\
&= \frac{f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})}{\int f(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi|\boldsymbol{D})\,d\phi} \\
&= \frac{\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}f(\phi|\boldsymbol{\zeta})}{\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}f(\phi|\boldsymbol{\zeta})\,d\phi} \\
&\propto \left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}f(\phi|\boldsymbol{\zeta}).
\end{aligned}
$$

Secondly, we fit the historical data $\boldsymbol{D}_0$ with Model (4.6) to draw posterior samples of $\boldsymbol{\beta'}$ from $g(\boldsymbol{\beta'}|\boldsymbol{D}_0)$. The joint posterior distribution of $\boldsymbol{\beta'}$,β_0' and ϕ' is

$$
\begin{aligned}
&g(\boldsymbol{\beta'}, \beta_0', \phi' | \boldsymbol{D}_0) = \frac{g(\boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{D}_0)}{g(\boldsymbol{D}_0)} \\
&= \frac{g(\boldsymbol{Z}|\boldsymbol{\beta'}, \beta_0', \phi') g(\boldsymbol{\beta'}) g(\beta_0') g(\phi'|\boldsymbol{\zeta'})}{\int \cdots \int g(\boldsymbol{Z}|\boldsymbol{\beta'}, \beta_0', \phi') g(\boldsymbol{\beta'}) g(\beta_0') g(\phi'|\boldsymbol{\zeta'})\, d\boldsymbol{\beta'}\, d\beta_0'\, d\phi'} \\
&\propto g(\boldsymbol{Z}|\boldsymbol{\beta'}, \beta_0', \phi') g(\boldsymbol{\beta'}) g(\beta_0') g(\phi'|\boldsymbol{\zeta'}) \\
&= \Big\{ \prod_{l=1}^{P-1} \int \cdots \int \exp\{ \frac{Z_l(\beta_0' + \boldsymbol{X}_l'^T \boldsymbol{\beta'}) - N_l d(\beta_0' + \boldsymbol{X}_l'^T \boldsymbol{\beta'})}{a(\phi')} \} \\
&\exp\{ \sum_{i=1}^{N_l - 1} c(Z_{il}, \phi') + c(Z_l - \sum_{i=1}^{N_l - 1} Z_{il}, \phi') \}\, dZ_{1l} \ldots\, dZ_{N_l - 1 l} \Big\} \\
&\Big\{ \prod_{p=1}^{P-1} \sqrt{\frac{\tau_{\beta_p'}}{2\pi}} e^{-\frac{\tau_{\beta_p'}}{2}(\beta_p' - \mu_{\beta_p'})^2} \Big\} \Big\{ \sqrt{\frac{\tau_{\beta_0'}}{2\pi}} e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0' - \mu_{\beta_0'})^2} \Big\}\, g(\phi'|\boldsymbol{\zeta'}) \\
&\propto \Big\{ \prod_{l=1}^{P-1} \int \cdots \int \exp\{ \frac{Z_l(\beta_0' + \boldsymbol{X}_l'^T \boldsymbol{\beta'}) - N_l d(\beta_0' + \boldsymbol{X}_l'^T \boldsymbol{\beta'})}{a(\phi')} \} \\
&\exp\{ \sum_{i=1}^{N_l - 1} c(Z_{il}, \phi') + c(Z_l - \sum_{i=1}^{N_l - 1} Z_{il}, \phi') \}\, dZ_{1l} \ldots\, dZ_{N_l - 1 l} \Big\} \\
&\Big\{ \prod_{p=1}^{P-1} e^{-\frac{\tau_{\beta_p'}}{2}(\beta_p' - \mu_{\beta_p'})^2} \Big\} \Big\{ e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0' - \mu_{\beta_0'})^2} \Big\}\, g(\phi'|\boldsymbol{\zeta'}),
\end{aligned}
$$

where $\boldsymbol{Z} = (Z_0, Z_1, \ldots, Z_{P-1})$. Therefore, the full conditionals of $\boldsymbol{\beta'}$, β_0' and ϕ' are

$$
\begin{aligned}
&g(\boldsymbol{\beta'}|\boldsymbol{D}_0, \beta_0', \phi') = \frac{g(\boldsymbol{\beta'}, \beta_0', \phi'|\boldsymbol{D}_0)}{g(\beta_0', \phi'|\boldsymbol{D}_0)} \\
&= \frac{g(\boldsymbol{\beta'}, \beta_0', \phi'|\boldsymbol{D}_0)}{\int \cdots \int g(\boldsymbol{\beta'}, \beta_0', \phi'|\boldsymbol{D}_0)\, d\boldsymbol{\beta'}} \\
&\propto \Big\{ \prod_{l=1}^{P-1} \int \cdots \int \exp\{ \frac{Z_l(\beta_0' + \boldsymbol{X}_l'^T \boldsymbol{\beta'}) - N_l d(\beta_0' + \boldsymbol{X}_l'^T \boldsymbol{\beta'})}{a(\phi')} \} \\
&\exp\{ \sum_{i=1}^{N_l - 1} c(Z_{il}, \phi') + c(Z_l - \sum_{i=1}^{N_l - 1} Z_{il}, \phi') \}\, dZ_{1l} \ldots\, dZ_{N_l - 1 l} \Big\} \Big\{ \prod_{p=1}^{P-1} e^{-\frac{\tau_{\beta_p'}}{2}(\beta_p' - \mu_{\beta_p'})^2} \Big\},
\end{aligned}
$$

$$
\begin{aligned}
&g(\beta_0'|\boldsymbol{D}_0, \boldsymbol{\beta'}, \phi') = \frac{g(\boldsymbol{\beta'}, \beta_0', \phi'|\boldsymbol{D}_0)}{g(\boldsymbol{\beta'}, \phi'|\boldsymbol{D}_0)} \\
&= \frac{g(\boldsymbol{\beta'}, \beta_0', \phi'|\boldsymbol{D}_0)}{\int g(\boldsymbol{\beta'}, \beta_0', \phi'|\boldsymbol{D}_0)\, d\beta_0'} \\
&\propto \Big\{ \prod_{l=1}^{P-1} \int \cdots \int \exp\{\frac{Z_l(\beta_0' + \boldsymbol{X}_l'^T\boldsymbol{\beta'}) - N_l d(\beta_0' + \boldsymbol{X}_l'^T\boldsymbol{\beta'})}{a(\phi')}\} \\
&\exp\{\sum_{i=1}^{N_l-1} c(Z_{il}, \phi') + c(Z_l - \sum_{i=1}^{N_l-1} Z_{il}, \phi')\}\, dZ_{1l} \ldots dZ_{N_l-1l} \Big\} \left\{ e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0' - \mu_{\beta_0'})^2} \right\},
\end{aligned}
$$

$$
\begin{aligned}
&g(\phi'|\boldsymbol{D}_0, \boldsymbol{\beta'}, \beta_0') = \frac{g(\boldsymbol{\beta'}, \beta_0', \phi'|\boldsymbol{D}_0)}{g(\boldsymbol{\beta'}, \beta_0'|\boldsymbol{D}_0)} \\
&= \frac{g(\boldsymbol{\beta'}, \beta_0', \phi'|\boldsymbol{D}_0)}{\int g(\boldsymbol{\beta'}, \beta_0', \phi'|\boldsymbol{D}_0)\, d\phi'} \\
&\propto \Big\{ \prod_{l=1}^{P-1} \int \cdots \int \exp\{\frac{Z_l(\beta_0' + \boldsymbol{X}_l'^T\boldsymbol{\beta'}) - N_l d(\beta_0' + \boldsymbol{X}_l'^T\boldsymbol{\beta'})}{a(\phi')}\} \\
&\exp\{\sum_{i=1}^{N_l-1} c(Z_{il}, \phi') + c(Z_l - \sum_{i=1}^{N_l-1} Z_{il}, \phi')\}\, dZ_{1l} \ldots dZ_{N_l-1l} \Big\} g(\phi'|\boldsymbol{\zeta'}).
\end{aligned}
$$

Finally, we perform a power prior analysis (4.12) with both data $\boldsymbol{D}$ and $\boldsymbol{D}_0$ to draw posterior samples of $\boldsymbol{\beta}$ from $h(\boldsymbol{\beta}|\boldsymbol{D}, \boldsymbol{D}_0, a)$. The augmented posterior distribution of $\boldsymbol{\beta}$,β_0,$\boldsymbol{b}$,τ_b,ϕ, $\boldsymbol{\beta'}$,β_0',ϕ',$\boldsymbol{\rho}$,$\boldsymbol{\mu}$ and $\boldsymbol{\tau}$ is

$$
\begin{aligned}
&h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho} | \boldsymbol{D}, \boldsymbol{D}_0, a) \\
&= \frac{h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}, \boldsymbol{D}, \boldsymbol{D}_0, a)}{h(\boldsymbol{D}, \boldsymbol{D}_0, a)} \\
&\propto L(\beta_0, \boldsymbol{\beta} | \boldsymbol{D}, \boldsymbol{b}) f(\beta_0) f(\boldsymbol{b} | \tau_b) f(\tau_b) f(\phi | \boldsymbol{\zeta}) (L(\beta_0', \boldsymbol{\beta'} | \boldsymbol{D}_0)^a g(\beta_0') g(\phi' | \boldsymbol{\zeta'}) \\
&\prod_{p=1}^{P-1} \left\{ h_p(\boldsymbol{\beta_p} | \boldsymbol{m}_p, \boldsymbol{\Lambda}_p) h_p(\mu_p | a_p, R_p) h_p(\tau_p | \kappa_p, \nu_p) h_p(\rho_p | c_p, d_p) \right\} \\
&= \left\{ \prod_{i=1}^{I} \prod_{j=1}^{J_i} \exp\{ \frac{Y_{ij}(\beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta} + b_i) - d(\beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta} + b_i)}{a(\phi)} + c(Y_{ij}, \phi) \} \right\} \left\{ \sqrt{\frac{\tau_{\beta_0}}{2\pi}} e^{-\frac{\tau_{\beta_0}}{2}(\beta_0 - \mu_{\beta_0})^2} \right\} \\
&\left\{ \prod_{i=1}^{I} \sqrt{\frac{\tau_b}{2\pi}} e^{-\frac{1}{2} b_i^2 \tau_b} \right\} \left\{ \frac{\nu_b^{\kappa_b}}{\Gamma(\kappa_b)} \tau_b^{\kappa_b - 1} e^{-\nu_b \tau_b} \right\} f(\phi | \boldsymbol{\zeta}) \\
&\left\{ \prod_{l=1}^{P-1} \int \cdots \int \exp\{ \frac{Z_l(\beta_0' + \boldsymbol{X}_l'^T \boldsymbol{\beta'}) - N_l d(\beta_0' + \boldsymbol{X}_l'^T \boldsymbol{\beta'})}{a(\phi')} \} \right. \\
&\left. \exp\{ \sum_{i=1}^{N_l - 1} c(Z_{il}, \phi') + c(Z_l - \sum_{i=1}^{N_l - 1} Z_{il}, \phi') \} \, dZ_{1l} \ldots \, dZ_{N_l - 1 l} \right\}^a \left\{ \sqrt{\frac{\tau_{\beta_0'}}{2\pi}} e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0' - \mu_{\beta_0'})^2} \right\} g(\phi' | \boldsymbol{\zeta'}) \\
&\prod_{p=1}^{P-1} \left\{ \frac{1}{2\pi} |\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p} - \boldsymbol{m_p})' \boldsymbol{\Lambda_p} (\boldsymbol{\beta_p} - \boldsymbol{m_p})}{2}} \sqrt{\frac{R_p}{2\pi}} e^{-\frac{R_p}{2}(\mu_p - a_p)^2} \frac{\nu_p^{\kappa_p}}{\Gamma(\kappa_p)} \tau_p^{\kappa_p - 1} e^{-\nu_p \tau_p} \frac{\rho_p^{c_p - 1} (1 - \rho_p)^{d_p - 1}}{B(c_p, d_p)} \right\} \\
&\propto \left\{ \prod_{i=1}^{I} \prod_{j=1}^{J_i} \exp\{ \frac{Y_{ij}(\beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta} + b_i) - d(\beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta} + b_i)}{a(\phi)} + c(Y_{ij}, \phi) \} \right\} \left\{ e^{-\frac{\tau_{\beta_0}}{2}(\beta_0 - \mu_{\beta_0})^2} \right\} \\
&\left\{ \prod_{i=1}^{I} \sqrt{\tau_b} e^{-\frac{1}{2} b_i^2 \tau_b} \right\} \left\{ \tau_b^{\kappa_b - 1} e^{-\nu_b \tau_b} \right\} f(\phi | \boldsymbol{\zeta}) \left\{ \prod_{l=1}^{P-1} \int \cdots \int \exp\{ \frac{Z_l(\beta_0' + \boldsymbol{X}_l'^T \boldsymbol{\beta'}) - N_l d(\beta_0' + \boldsymbol{X}_l'^T \boldsymbol{\beta'})}{a(\phi')} \} \right. \\
&\left. \exp\{ \sum_{i=1}^{N_l - 1} c(Z_{il}, \phi') + c(Z_l - \sum_{i=1}^{N_l - 1} Z_{il}, \phi') \} \, dZ_{1l} \ldots \, dZ_{N_l - 1 l} \right\}^a \left\{ e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0' - \mu_{\beta_0'})^2} \right\} g(\phi' | \boldsymbol{\zeta'}) \\
&\prod_{p=1}^{P-1} \left\{ |\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p} - \boldsymbol{m_p})' \boldsymbol{\Lambda_p} (\boldsymbol{\beta_p} - \boldsymbol{m_p})}{2}} e^{-\frac{R_p}{2}(\mu_p - a_p)^2} \tau_p^{\kappa_p - 1} e^{-\nu_p \tau_p} \rho_p^{c_p - 1} (1 - \rho_p)^{d_p - 1} \right\},
\end{aligned}
$$

so the full conditionals of $\boldsymbol{\beta}$,β_0,$\boldsymbol{b}$,τ_b,ϕ,$\boldsymbol{\beta'}$,β_0',ϕ',ρ_p, μ_p, τ_p, $p = 1, 2, \ldots, P-1$ are

$$
\begin{aligned}
&h(\boldsymbol{\beta}|\boldsymbol{D},\boldsymbol{D}_0,a,\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho})\\
&=h(\boldsymbol{\beta}|\boldsymbol{D},\beta_0,\boldsymbol{b},\phi,\boldsymbol{\beta'},\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho})\\
&=\frac{h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\boldsymbol{\beta}}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\\
&\prod_{p=1}^{P-1}\left\{|\boldsymbol{\Lambda_p}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_p}-\boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p}-\boldsymbol{m_p})}{2}}\right\}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\\
&\prod_{p=1}^{P-1}\left\{e^{-\frac{\tau_p}{2(1-\rho_p^2)}\{(\beta_p-\mu_p)^2-2\rho_p(\beta_p-\mu_p)(\beta_p'-\mu_p)\}}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&h(\beta_0|\boldsymbol{D},\boldsymbol{D}_0,a,\boldsymbol{\beta},\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho})\\
&=h(\beta_0|\boldsymbol{D},\boldsymbol{\beta},\boldsymbol{b},\phi)\\
&=\frac{h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\beta_0}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{e^{-\frac{\tau_{\beta_0}}{2}(\beta_0-\mu_{\beta_0})^2}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&h(\phi|\boldsymbol{D},\boldsymbol{D}_0,a,\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho})\\
&=h(\phi|\boldsymbol{D},\boldsymbol{\beta},\beta_0,\boldsymbol{b})\\
&=\frac{h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\phi}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}f(\phi|\boldsymbol{\zeta}),
\end{aligned}
$$

$$
\begin{aligned}
&h(\boldsymbol{b}|\boldsymbol{D},\boldsymbol{D}_0,a,\boldsymbol{\beta},\beta_0,\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho})\\
&=h(\boldsymbol{b}|\boldsymbol{D},\boldsymbol{\beta},\beta_0,\tau_b,\phi)\\
&=\frac{h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\boldsymbol{b}}\\
&\propto\left\{\prod_{i=1}^{I}\prod_{j=1}^{J_i}\exp\{\frac{Y_{ij}(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)-d(\beta_0+\boldsymbol{X}_i^T\boldsymbol{\beta}+b_i)}{a(\phi)}+c(Y_{ij},\phi)\}\right\}\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&h(\tau_b|\boldsymbol{D},\boldsymbol{D}_0,a,\boldsymbol{\beta},\beta_0,\boldsymbol{b},\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho})\\
&=h(\tau_b|\boldsymbol{b})\\
&=\frac{h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\beta,\alpha,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\tau_b}\\
&=\frac{\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}}{\int\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}\,d\tau_b}\\
&\propto\left\{\prod_{i=1}^{I}\sqrt{\tau_b}e^{-\frac{1}{2}b_i^2\tau_b}\right\}\left\{\tau_b^{\kappa_b-1}e^{-\nu_b\tau_b}\right\}\\
&\sim\text{Gamma}(\kappa_b+\frac{I}{2},\nu_b+\frac{\sum_{i=1}^{I}b_i^2}{2}),
\end{aligned}
$$

$$
\begin{aligned}
&h(\boldsymbol{\beta'}|\boldsymbol{D},\boldsymbol{D}_0,a,\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho})\\
&=h(\boldsymbol{\beta'}|\boldsymbol{D}_0,a,\boldsymbol{\beta},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho})\\
&=\frac{h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)}{\int h(\boldsymbol{\beta},\beta_0,\boldsymbol{b},\tau_b,\phi,\boldsymbol{\beta'},\beta_0',\phi',\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\rho}|\boldsymbol{D},\boldsymbol{D}_0,a)\,d\boldsymbol{\beta'}}\\
&\propto\left\{\prod_{l=1}^{P-1}\int\cdots\int\exp\{\frac{Z_l(\beta_0'+\boldsymbol{X}_l'^T\boldsymbol{\beta'})-N_l d(\beta_0'+\boldsymbol{X}_l'^T\boldsymbol{\beta'})}{a(\phi')}\}\right.\\
&\left.\exp\{\sum_{i=1}^{N_l-1}c(Z_{il},\phi')+c(Z_l-\sum_{i=1}^{N_l-1}Z_{il},\phi')\}\,dZ_{1l}\ldots\,dZ_{N_l-1l}\right\}^a\prod_{p=1}^{P-1}\left\{|\boldsymbol{\Lambda_p}|^{\frac{1}{2}}e^{-\frac{(\boldsymbol{\beta_p}-\boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p}-\boldsymbol{m_p})}{2}}\right\}\\
&\propto\left\{\prod_{l=1}^{P-1}\int\cdots\int\exp\{\frac{Z_l(\beta_0'+\boldsymbol{X}_l'^T\boldsymbol{\beta'})-N_l d(\beta_0'+\boldsymbol{X}_l'^T\boldsymbol{\beta'})}{a(\phi')}\}\right.\\
&\left.\exp\{\sum_{i=1}^{N_l-1}c(Z_{il},\phi')+c(Z_l-\sum_{i=1}^{N_l-1}Z_{il},\phi')\}\,dZ_{1l}\ldots\,dZ_{N_l-1l}\right\}^a\\
&\prod_{p=1}^{P-1}\left\{e^{-\frac{\tau_p}{2(1-\rho_p^2)}\{(\beta_p'-\mu_p)^2-2\rho_p(\beta_p-\mu_p)(\beta_p'-\mu_p)\}}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&h(\beta_0'|\boldsymbol{D}, \boldsymbol{D}_0, a, \boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}) \\
&= h(\beta_0'|\boldsymbol{D}_0, a, \boldsymbol{\beta'}, \phi') \\
&= \frac{h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}|\boldsymbol{D}, \boldsymbol{D}_0, a)}{\int h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}|\boldsymbol{D}, \boldsymbol{D}_0, a)\, d\beta_0'} \\
&\propto \left\{ \prod_{l=1}^{P-1} \int \cdots \int \exp\{\frac{Z_l(\beta_0' + \boldsymbol{X}_l'^T\boldsymbol{\beta'}) - N_l d(\beta_0' + \boldsymbol{X}_l'^T\boldsymbol{\beta'})}{a(\phi')}\} \right. \\
&\left. \exp\{\sum_{i=1}^{N_l-1} c(Z_{il}, \phi') + c(Z_l - \sum_{i=1}^{N_l-1} Z_{il}, \phi')\}\, dZ_{1l} \ldots dZ_{N_l-1l} \right\}^a \left\{ e^{-\frac{\tau_{\beta_0'}}{2}(\beta_0' - \mu_{\beta_0'})^2} \right\},
\end{aligned}
$$

$$
\begin{aligned}
&h(\phi'|\boldsymbol{D}, \boldsymbol{D}_0, a, \boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}) \\
&= h(\phi'|\boldsymbol{D}_0, a, \boldsymbol{\beta'}, \beta_0') \\
&= \frac{h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}|\boldsymbol{D}, \boldsymbol{D}_0, a)}{\int h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}|\boldsymbol{D}, \boldsymbol{D}_0, a)\, d\phi'} \\
&\propto \left\{ \prod_{l=1}^{P-1} \int \cdots \int \exp\{\frac{Z_l(\beta_0' + \boldsymbol{X}_l'^T\boldsymbol{\beta'}) - N_l d(\beta_0' + \boldsymbol{X}_l'^T\boldsymbol{\beta'})}{a(\phi')}\} \right. \\
&\left. \exp\{\sum_{i=1}^{N_l-1} c(Z_{il}, \phi') + c(Z_l - \sum_{i=1}^{N_l-1} Z_{il}, \phi')\}\, dZ_{1l} \ldots dZ_{N_l-1l} \right\}^a g(\phi'|\boldsymbol{\zeta'}),
\end{aligned}
$$

$$
\begin{aligned}
&h(\rho_p|\boldsymbol{D}, \boldsymbol{D}_0, a, \boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}_{-p}) \\
&= h(\rho_p|\beta_p, \beta_p', \mu_p, \tau_p) \\
&= \frac{h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}|\boldsymbol{D}, \boldsymbol{D}_0, a)}{\int h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}|\boldsymbol{D}, \boldsymbol{D}_0, a)\, d\rho_p} \\
&= \frac{\left\{ |\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p} - \boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p} - \boldsymbol{m_p})}{2}} \right\} \left\{ \rho_p^{c_p - 1}(1 - \rho_p)^{d_p - 1} \right\}}{\int \left\{ |\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p} - \boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p} - \boldsymbol{m_p})}{2}} \right\} \left\{ \rho_p^{c_p - 1}(1 - \rho_p)^{d_p - 1} \right\} d\rho_p} \\
&\propto \left\{ |\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p} - \boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p} - \boldsymbol{m_p})}{2}} \right\} \left\{ \rho_p^{c_p - 1}(1 - \rho_p)^{d_p - 1} \right\} \\
&\propto \rho_p^{c_p - 1}(1 - \rho_p)^{d_p - 1} e^{-\tau_p \frac{(\beta_p - \mu_p)^2 - 2\rho_p(\beta_p - \mu_p)(\beta_p' - \mu_p) + (\beta_p' - \mu_p)^2}{2(1 - \rho_p^2)}}.
\end{aligned}
$$

$$
\begin{aligned}
&h(\mu_p|\boldsymbol{D}, \boldsymbol{D}_0, a, \boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}_{-p}, \boldsymbol{\tau}, \boldsymbol{\rho}) \\
&= h(\mu_p|\beta_p, \beta_p', \tau_p, \rho_p) \\
&= \frac{h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}|\boldsymbol{D}, \boldsymbol{D}_0, a)}{\int h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}|\boldsymbol{D}, \boldsymbol{D}_0, a)\, d\mu_p} \\
&= \frac{\left\{|\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p}-\boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p}-\boldsymbol{m_p})}{2}}\right\}\left\{e^{-\frac{R_p}{2}(\mu_p-a_p)^2}\right\}}{\int \left\{|\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p}-\boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p}-\boldsymbol{m_p})}{2}}\right\}\left\{e^{-\frac{R_p}{2}(\mu_p-a_p)^2}\right\} d\mu_p} \\
&\propto \left\{|\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p}-\boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p}-\boldsymbol{m_p})}{2}}\right\}\left\{e^{-\frac{R_p}{2}(\mu_p-a_p)^2}\right\} \\
&\propto e^{-\frac{1}{2}\{(\boldsymbol{\beta_p}-\boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p}-\boldsymbol{m_p})+R_p(\mu_p-a_p)^2\}} \\
&\propto e^{-\frac{1}{2}(\frac{2\tau_p}{1+\rho_p}+R_p)(\mu_p-\frac{\frac{\tau_p}{1+\rho_p}(\beta_p+\beta_p')+R_p a_p}{\frac{2\tau_p}{1+\rho_p}+R_p})^2} \\
&\sim \mathrm{N}(\frac{\frac{\tau_p}{1+\rho_p}(\beta_p+\beta_p')+R_p a_p}{\frac{2\tau_p}{1+\rho_p}+R_p}, \frac{2\tau_p}{1+\rho_p}+R_p),
\end{aligned}
$$

$$
\begin{aligned}
&h(\tau_p|\boldsymbol{D}, \boldsymbol{D}_0, a, \boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}_{-p}, \boldsymbol{\rho}) \\
&= h(\tau_p|\beta_p, \beta_p', \mu_p, \rho_p) \\
&= \frac{h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}|\boldsymbol{D}, \boldsymbol{D}_0, a)}{\int h(\boldsymbol{\beta}, \beta_0, \boldsymbol{b}, \tau_b, \phi, \boldsymbol{\beta'}, \beta_0', \phi', \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}|\boldsymbol{D}, \boldsymbol{D}_0, a)\, d\tau_p} \\
&= \frac{\left\{|\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p}-\boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p}-\boldsymbol{m_p})}{2}}\right\}\left\{\tau_p^{\kappa_p-1} e^{-\nu_p\tau_p}\right\}}{\int \left\{|\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p}-\boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p}-\boldsymbol{m_p})}{2}}\right\}\left\{\tau_p^{\kappa_p-1} e^{-\nu_p\tau_p}\right\} d\tau_p} \\
&\propto \left\{|\boldsymbol{\Lambda_p}|^{\frac{1}{2}} e^{-\frac{(\boldsymbol{\beta_p}-\boldsymbol{m_p})'\boldsymbol{\Lambda_p}(\boldsymbol{\beta_p}-\boldsymbol{m_p})}{2}}\right\}\left\{\tau_p^{\kappa_p-1} e^{-\nu_p\tau_p}\right\} \\
&= \tau_p^{\kappa_p+\frac{1}{2}-1} e^{-\tau_p(\nu_p+\frac{(\beta_p-\mu_p)^2-2\rho_p(\beta_p-\mu_p)(\beta_p'-\mu_p)+(\beta_p'-\mu_p)^2}{2(1-\rho_p^2)})} \\
&\sim \mathrm{Gamma}(\kappa_p+\frac{1}{2}, \nu_p+\frac{(\beta_p-\mu_p)^2-2\rho_p(\beta_p-\mu_p)(\beta_p'-\mu_p)+(\beta_p'-\mu_p)^2}{2(1-\rho_p^2)}),
\end{aligned}
$$

The Metropolis Hastings Component of the Hybrid Algorithm

When the Gibbs algorithm is used to draw posterior samples of $\boldsymbol{\beta}$ from $f(\boldsymbol{\beta}|\boldsymbol{D})$, $\boldsymbol{\beta'}$ from $g(\boldsymbol{\beta'}|\boldsymbol{D}_0)$ or $\boldsymbol{\beta}$ from $h(\boldsymbol{\beta}|\boldsymbol{D}, \boldsymbol{D}_0, a)$, not all of the full conditionals have closed forms. So we construct a Markov chain with the full conditional as the target

distribution to get the corresponding posterior samples. This is the MH part of the hybrid algorithm. We develop a random walk MH algorithm with a t proposal density. Since the algorithm can be implemented for the parameters in the same way, we write the MH algorithm in a more general notation.

We write the parameter of interest and the nuisance parameters as $\boldsymbol{\theta}$ and $\boldsymbol{\theta'}$, respectively, and write the data as $\boldsymbol{D}_g$. Assuming that the full conditional is $k(\boldsymbol{\theta}|\boldsymbol{D}_g, \boldsymbol{\theta'})$, we note that the MH part of the hybrid algorithm for $\boldsymbol{\theta}$ can be carried out as follows.

1. Start with $\boldsymbol{\theta}^0$ and set $m = 1$.
2. Sample $\boldsymbol{\theta}^* \sim t(\boldsymbol{\theta}^{m-1}, \nu_t)$, and accept $\boldsymbol{\theta}^*$ with probability

$$\min\{1, \frac{k(\boldsymbol{\theta}^*|\boldsymbol{D}_g, \boldsymbol{\theta'})}{k(\boldsymbol{\theta}^{m-1}|\boldsymbol{D}_g, \boldsymbol{\theta'})}\},$$

   and set $\boldsymbol{\theta}^m = \boldsymbol{\theta}^*$. Otherwise, set $\boldsymbol{\theta}^m = \boldsymbol{\theta}^{m-1}$.
3. Increase m by 1 and return to step 2.

Here $t(\boldsymbol{\theta}^{m-1}, \nu_t)$ is the univariate or multivariate t distribution with $\boldsymbol{\theta}^{m-1}$ and ν_t as the location and degree of freedom, depending on whether $\boldsymbol{\theta}$ is a random variable or vector. With this hybrid algorithm, the posterior samples of $\boldsymbol{\theta}$ are drawn from its posterior $k(\boldsymbol{\theta}|\boldsymbol{D}_g)$.

4.5 Simulation Studies

4.5.1 Simulation Settings and Data Generation

We conducted a simulation study to examine the performance of this power prior method in cluster randomized trials with (1) two arms in normal data; (2) two arms

in count data; and (3) three arms in binary data. The trials were assumed to be balanced in treatment group size. For two-arm trials with normal data, the model for generating the current data $\boldsymbol{D} = \{(X_{1i}, Y_{ij}) : i = 1, 2, \ldots, I, j = 1, 2, \ldots, J\}$ was $Y_{ij} \sim \text{Normal}(\mu_i, \tau_1)$, $\mu_i = \beta_0 + \beta_1 X_{1i} + b_i$,$b_i \sim \text{Normal}(0, \tau_b)$, and the model for generating the historical data $\boldsymbol{D_0} = \{(X'_{1l}, Z_l) : l = 0, 1\}$ was $Z_l \sim \text{Normal}(N\mu_l, \frac{\tau_2}{N})$, $\mu_l = \beta'_0 + \beta'_1 X'_{1l}$, where half of X_{1i} is 0 for placebo arm and the other half is 1 for treatment arm. $X'_{1l} = 0$ when $l = 0$ for placebo arm and it is equal to 1 when $l = 1$ for treatment arm.

For two-arm trials with count data, the current trial data model, however, was $Y_{ij} \sim \text{Poisson}(\lambda_i)$, $\log(\lambda_i) = \beta_0 + \beta_1 X_{1i} + b_i$, $b_i \sim \text{Normal}(0, \tau_b)$, and the historical trial data model was $Z_l \sim \text{Poisson}(N\lambda_l)$,$\log(\lambda_l) = \beta'_0 + \beta'_1 X'_{1l}$. X_{1i} and X'_{1l} take the same values as those in two arm trials with normal data. For three-arm trials with binary data, the current trial data $\boldsymbol{D} = \{(X_{1i}, X_{2i}, Y_{ij}) : i = 1, 2, \ldots, I, j = 1, 2, \ldots, J\}$ was generated by the model $Y_{ij} \sim \text{Bernoulli}(p_i)$, $\text{logit}(p_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + b_i$, $b_i \sim \text{Normal}(0, \tau_b)$, and the historical trial data $\boldsymbol{D_0} = \{(X'_{1l}, X'_{2l}, Z_l) : l = 0, 1, 2\}$ was generated by the model $Z_l \sim \text{Binomial}(N, q_l)$, $\text{logit}(q_l) = \beta'_0 + \beta'_1 X'_{1l} + \beta'_2 X'_{2l}$. The first third of (X_{1i}, X_{2i}) is $(0, 0)$ for the placebo arm, the second third of (X_{1i}, X_{2i}) is $(1, 0)$ for the first treatment arm, and the remaining is $(0, 1)$ for the second treatment arm. Additionally, $(X'_{1l}, X'_{2l}) = (0, 0)$ when $l = 0$ for placebo arm, $(1, 0)$ when $l = 1$ for the first treatment arm and $(0, 1)$ when $l = 2$ for the second treatment arm.

In each of these three cases, simulation studies were carried out to investigate the effects of the treatment effect (β_1, β'_1) in two arm trials and $(\beta_1, \beta_2, \beta'_1, \beta'_2)$ in three arm trials, the historical arm size N, the total cluster number I, cluster size J, the ICC ρ in current trial. For normal data, $\rho = \frac{\tau_b}{\tau_b + \tau_1}$, and for count data, the ICC

changes between different arms (Stryhn et al., 2006). Let the ICC be ρ_+ and ρ_- for treatment arm and placebo arm, respectively, then we could calculate them as

$$\rho_+ = \frac{\exp(2*(\beta_0+\beta_1+\frac{1}{\tau_b})) - \exp(2*(\beta_0+\beta_1)+\frac{1}{\tau_b})}{\exp(2*(\beta_0+\beta_1+\frac{1}{\tau_b})) - \exp(2*(\beta_0+\beta_1)+\frac{1}{\tau_b}) + \exp(\beta_0+\beta_1+\frac{1}{2\tau_b})}$$

and,

$$\rho_- = \frac{\exp(2*(\beta_0+\frac{1}{\tau_b})) - \exp(2*\beta_0+\frac{1}{\tau_b})}{\exp(2*(\beta_0+\frac{1}{\tau_b})) - \exp(2*\beta_0+\frac{1}{\tau_b}) + \exp(\beta_0+\frac{1}{2\tau_b})}.$$

For binary data, $\rho = \frac{1}{1+\frac{\pi^2}{3}\tau_b}$. For the purpose of illustration, we used the study of the effect of treatment effect (β_1, β_1') as an example in this section. The other results about the effect of N, I, J and ρ could be required from the authors once requested. The settings of the current trial and historical trial were given in Tables 4.1 and 4.2, respectively.

For each parameter setting in Tables 4.1 and 4.2, $1,000$ data replicates were simulated from the corresponding model, and each data replicate was analyzed with the power prior method under four different discounting parameters, that is, $0, 1, a^{sym}$ and a^{asym}. The posterior median, SD and the 95% HPD interval of β_1 were computed. Additionally, the posterior median and SD of β_2, the 95% HPD interval of $\beta_1 - \beta_2$ and the 97.5% HPD intervals of β_1 and β_2 were calculated as well for three-arm trials. The 95% HPD interval was used to test the null hypothesis $\mathrm{H}_0 : \mathrm{E}(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0) = 0$ or $\mathrm{H}_0 : \mathrm{E}(\beta_1 - \beta_2|\boldsymbol{D}, \boldsymbol{D}_0) = 0$, and the 97.5% HPD interval was used to test the null hypothesis $\mathrm{H}_0 : \mathrm{E}(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0) = \mathrm{E}(\beta_2|\boldsymbol{D}, \boldsymbol{D}_0) = 0$. By averaging the $1,000$ sets of analytical results, we could calculate the estimate of the posterior median, SD, relative bias, rMSE of β_1 (and β_2 and $\beta_1 - \beta_2$ for three-arm trials as well), and the average length, CP and RP of the HPD intervals. The RP was type I error rate

when $\beta_1 = 0$ and power otherwise for $\mathrm{H}_0 : \mathrm{E}(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0) = 0$, and $\beta_1 - \beta_2 = 0$ for $\mathrm{H}_0 : \mathrm{E}(\beta_1 - \beta_2|\boldsymbol{D}, \boldsymbol{D}_0) = 0$, and $\beta_1 = \beta_2 = 0$ for $\mathrm{H}_0 : \mathrm{E}(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0) = \mathrm{E}(\beta_2|\boldsymbol{D}, \boldsymbol{D}_0) = 0$. To calculate the KL divergence based discounting parameters, separate data analyses of current and historical trials were done under non-informative priors within Bayesian framework. The MH within Gibbs algorithm was used to complete the computation, and a Markov Chain of length $2,500$ was run with the first $1,500$ samples discarded and the thinning rate being 5. In the power prior data analysis, however, a Markov Chain of length $5,000$ was run with the first $2,500$ samples discarded and the thinning rate being 5. In both scenarios, a t proposal density was used, and its degree of freedom ν_t was set to be 3. The MCMC convergence was diagnosed with the trace plots. To end this section, we would point out that, in these three cases, we also developed similar simulation based algorithm in the design of cluster randomized trials as before, and simulation studies were done to determine the sample size. These results could be required from the authors once requested.

4.5.2 Simulation Results

For illustrative purpose, we reported the effect of the treatment effect (β_1, β_1') on the discounting fraction, SD, rMSE and type I error rate and power in the three different cases. All results were summarized in Figures $4.1 - 4.5$.

Figure 4.1 shows the results from the two-arm trials with normal data. From the animation plot of discounting fraction, we see that the estimate of the symmetric and asymmetric KL divergence based discounting fractions, a^{sym} and a^{asym}, would increase as the historical treatment effect β_1' got closer to the current one β_1. As a^{sym} and a^{asym} became larger, more historical information similar to the current one

were borrowed, and therefore, the SD and rMSE would become smaller, which was illustrated in the animation plots of SD and rMSE. The animation plot of power shows that it would increase as more similar historical information were borrowed and that it would decrease otherwise. The opposite trend was observed for type I error rate from its animation plot.

Figures $4.2 - 4.3$ demonstrate the same information about the discounting fraction, SD, rMSE and type I error rate and power for the null hypothesis H_0 : $\mathrm{E}(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0) = 0$ in two-arm trials with count data and three-arm trials with binary data. Furthermore, in three-arm trials with binary data, we fixed $\beta_2 = \beta_2' = \log(1) = 0$, and we investigated the effect of $\beta_1' - \beta_2' = \beta_1'$ on the SD and rMSE of $\beta_1 - \beta_2 = \beta_1$ and the type I error rate and power for the null hypothesis $\mathrm{H}_0 : \mathrm{E}(\beta_1 - \beta_2|\boldsymbol{D}, \boldsymbol{D}_0) = 0$ with Figure 4.4. The same message could be read from this Figure as those from Figures $4.1 - 4.3$. Finally, we used the 97.5% HPD intervals of β_1 and β_2 to do the simultaneous test of $\mathrm{H}_0 : \mathrm{E}(\beta_1|\boldsymbol{D}, \boldsymbol{D}_0) = \mathrm{E}(\beta_2|\boldsymbol{D}, \boldsymbol{D}_0) = 0$, and the animation plot of the type I error rate and power was given in Figure 4.5. As a by-product, the animation plots of SD and rMSE of β_2 were also given in Figure 4.5. From this figure, we see the SD and rMSE of β_2 decreased as more similar historical information was borrowed (see the animation plot of discounting fraction in Figure 4.3). As similar historical information was borrowed, the power would increase and the type I error rate would decrease.

4.6 Discussion

In this chapter, we extended our power prior approach to multi-arm cluster randomized trial with outcome from exponential family distributions, and we used well-

designed simulation studies to evaluate its performance in three different situations, including two-arm trial with normal outcome, two-arm trial with count outcome and three-arm trial with binary outcome. All simulation results showed that our power prior method was useful in the design and analysis of cluster randomized trial in these situations. After borrowing historical information with this power prior method, we could expect to increase the power by 10 percent under some conditions, and at the same time,we could control the type I error rate. This meant fewer samples would be needed to achieve a pre-specified power with the type I error rate controlled.

In the Simulation Studies section, we used the animation plots to present the results. This was a very useful way to visualize high dimensional data. Compared to traditional plots, the animation plots could help the readers recognize the key findings behind the results quickly and save space as well. This method was expected to have some values in applications, however, some of its users, such as trialists or health care practitioners, might have little statistical and programming knowledge. Therefore it was necessary to develop a computational tool for this method so that the users could use it to design and analyze cluster randomized trial conveniently. For example, they might only need to upload the data onto the computational tool and click certain buttons to get the results. This computational tool was expected to have three functionalities, including (1) analyzing a cluster randomized trial with a historical trial data; (2) designing a cluster randomized trial given a historical trial data; and (3) visualizing high dimensional data with animation plots. Since our program was developed with R, we intended to develop the computational tool as a R package, and submit it onto Github or CRAN for users to download and use it. This topic would be covered in the next chapter.

Table 4.1: Current trial setting with different types of data and numbers of arm

Data type	Arm number	I	J	β_0	β_1	β_2	τ_b	τ_1
Normal	2	10	10	-0.1	$0, 1, 1.5, \ldots, 3$	–	1	1
Count	2	20	20	-0.1	$\log(1), \log(1.2), \ldots, \log(2)$	–	4	–
Binary	3	30	20	-0.1	$\log(1), \log(4)$	$\log(1)$	1	–

Table 4.2: Historical trial setting with different types of data and numbers of arm

Data type	Arm number	N	β_0'	β_1'	β_2'	τ_2
Normal	2	200	0.1	$0, 1, 1.5, \ldots, 3$	–	1
Count	2	200	0.1	$\log(1), \log(1.2), \ldots, \log(2)$	–	–
Binary	3	200	0.1	$\log(1), \log(2), \ldots, \log(6)$	$\log(1)$	–

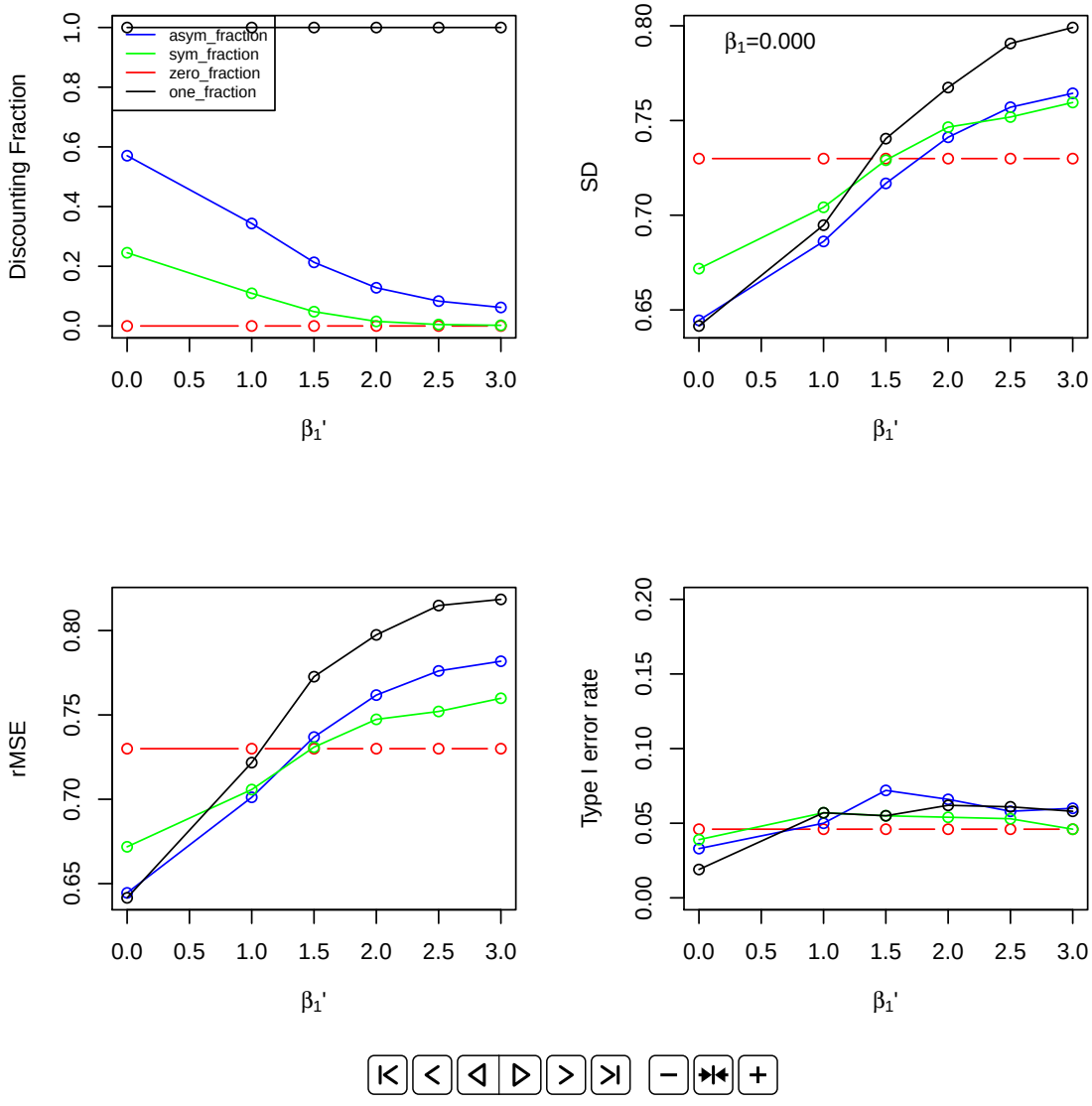

Figure 4.1: In two-arm trial with normal data, the effect of the historical treatment effect β_1' on the discounting fraction, SD, rMSE, type I error rate and power where the total number of cluster $I = 10$,the cluster size $J = 10$, and the ICC $\rho = 0.5$ in current trial and the arm size $N = 200$ in historical trial. Type I error rate and power correspond to $\beta_1 = 0$ and $\beta_1 = 1, 1.5, 2, 2.5, 3$.

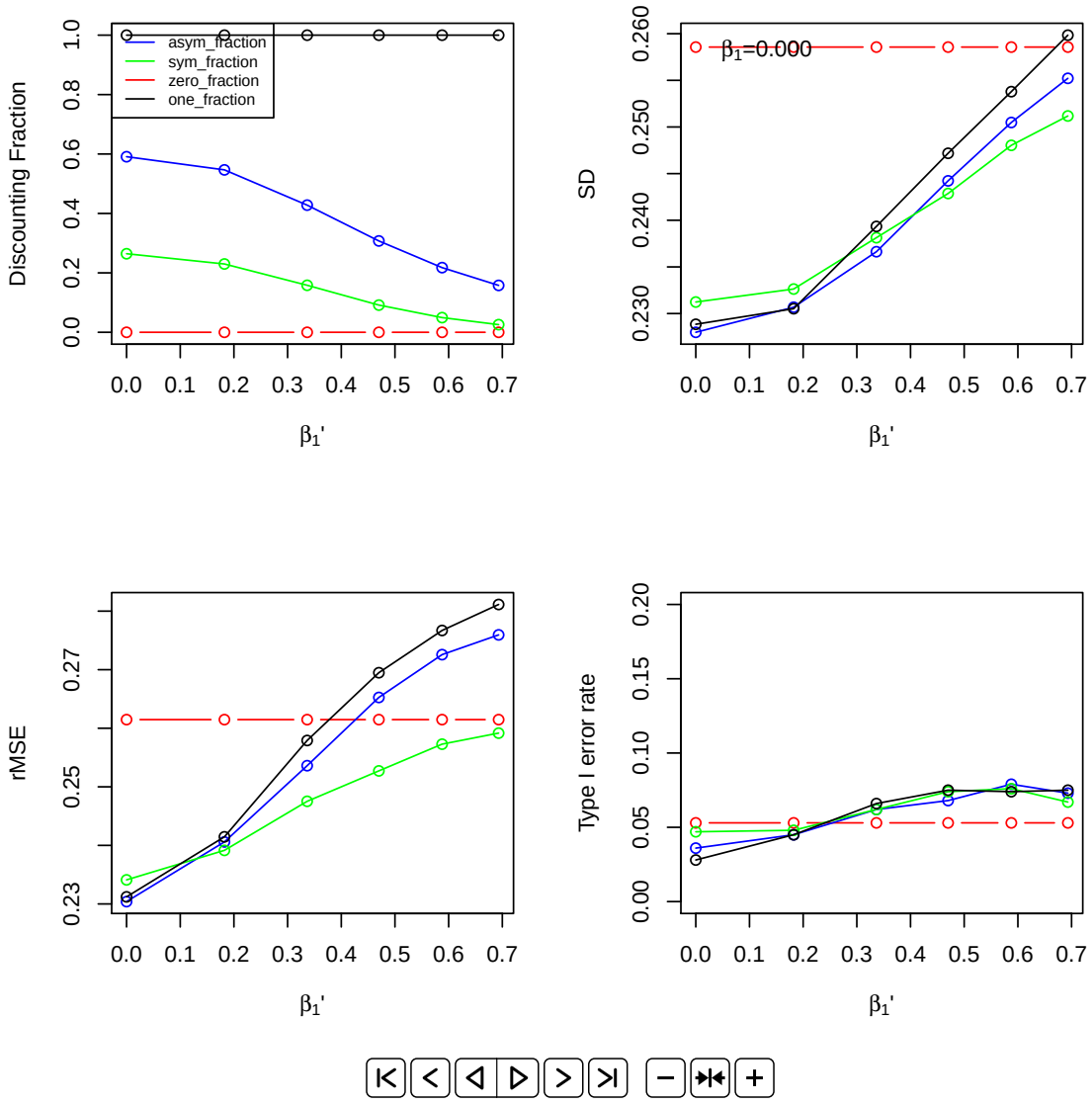

Figure 4.2: In two-arm trial with count data, the effect of the historical treatment effect β_1' on the discounting fraction, SD, rMSE, type I error rate and power where the total number of cluster $I = 20$,the cluster size $J = 20$ in current trial and the arm size $N = 200$ in historical trial. Type I error rate and power correspond to $\beta_1 = \log(1) = 0$ and $\beta_1 = \log(1.2), \log(1.4), \log(1.6), \log(1.8), \log(2) \approx 0.182, 0.336, 0.470, 0.588, 0.693$. The ICC are $\rho_+ = 0.226, 0.259, 0.290, 0.318,\ 0.344, 0.368$ for the treatment arm and $\rho_- = 0.226$ for the placebo arm where $\beta_1 = 0, 0.182, 0.336, 0.470, 0.588, 0.693$.

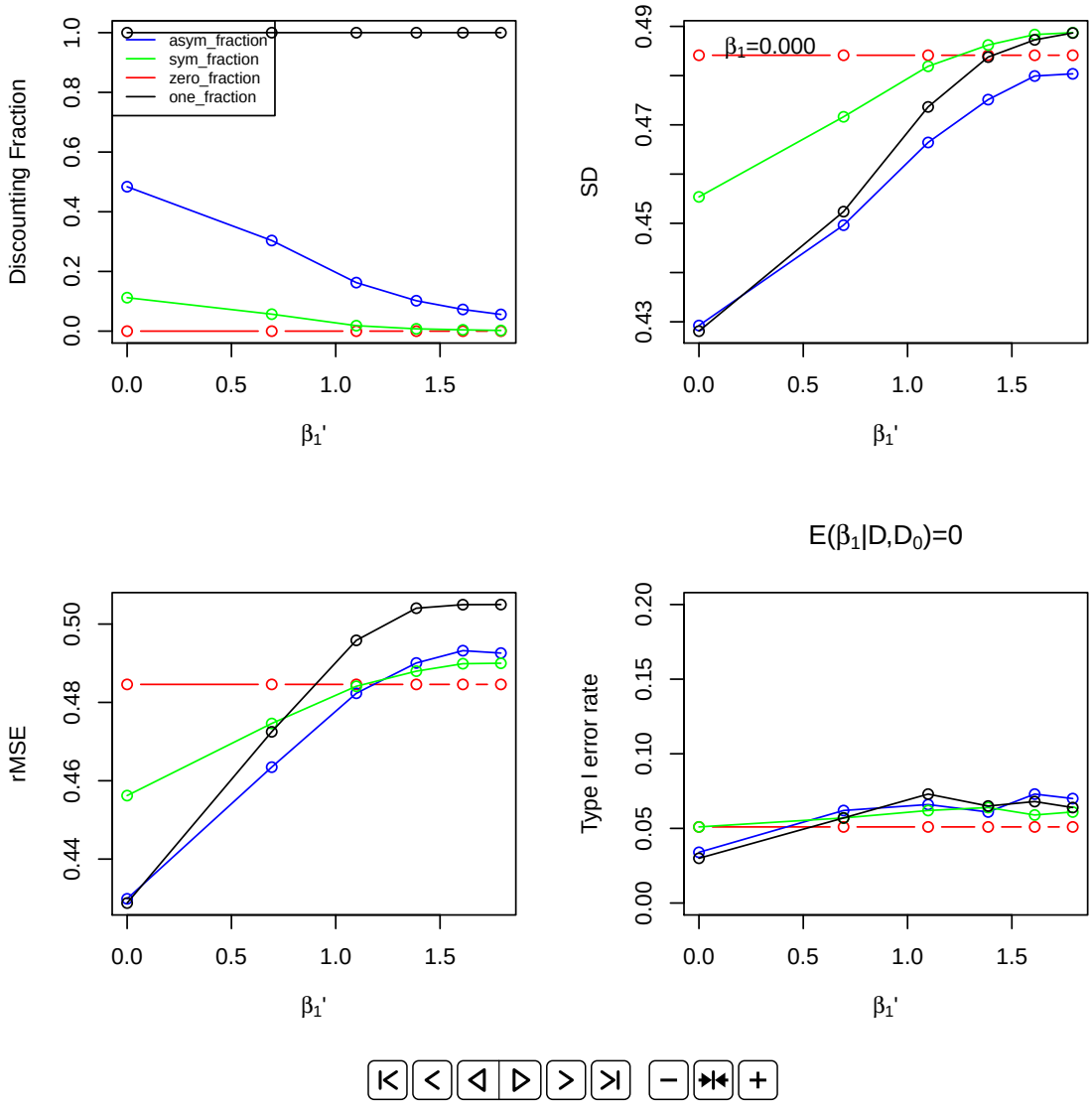

Figure 4.3: In three-arm trial with binary data, the effect of the historical treatment effect β_1' on the discounting fraction, SD and rMSE of β_1, type I error rate and power where the total number of cluster $I = 30$,the cluster size $J = 20$, and the ICC $\rho = 0.233$ in current trial and the arm size $N = 200$ in historical trial. Type I error rate and power correspond to $\beta_1 = \log(1) = 0$ and $\beta_1 = \log(4) \approx 1.386$.

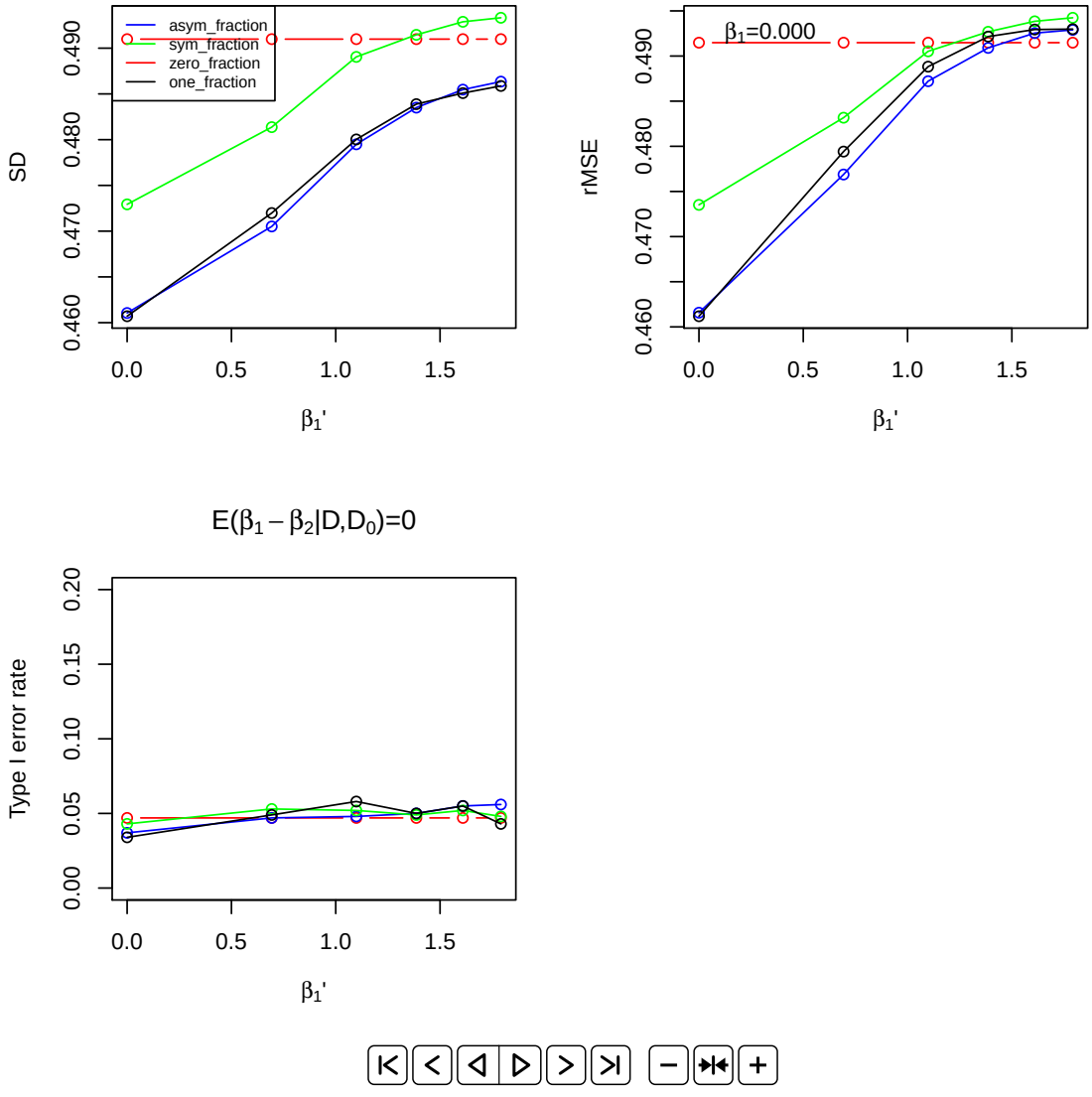

Figure 4.4: In three-arm trial with binary data, the effect of the historical treatment effect β_1' on the discounting fraction, SD and rMSE of $\beta_1 - \beta_2$, type I error rate and power where the total number of cluster $I = 30$,the cluster size $J = 20$, and the ICC $\rho = 0.233$ in current trial and the arm size $N = 200$ in historical trial. Type I error rate and power correspond to $\beta_1 = \log(1) = 0$ and $\beta_1 = \log(4) \approx 1.386$.

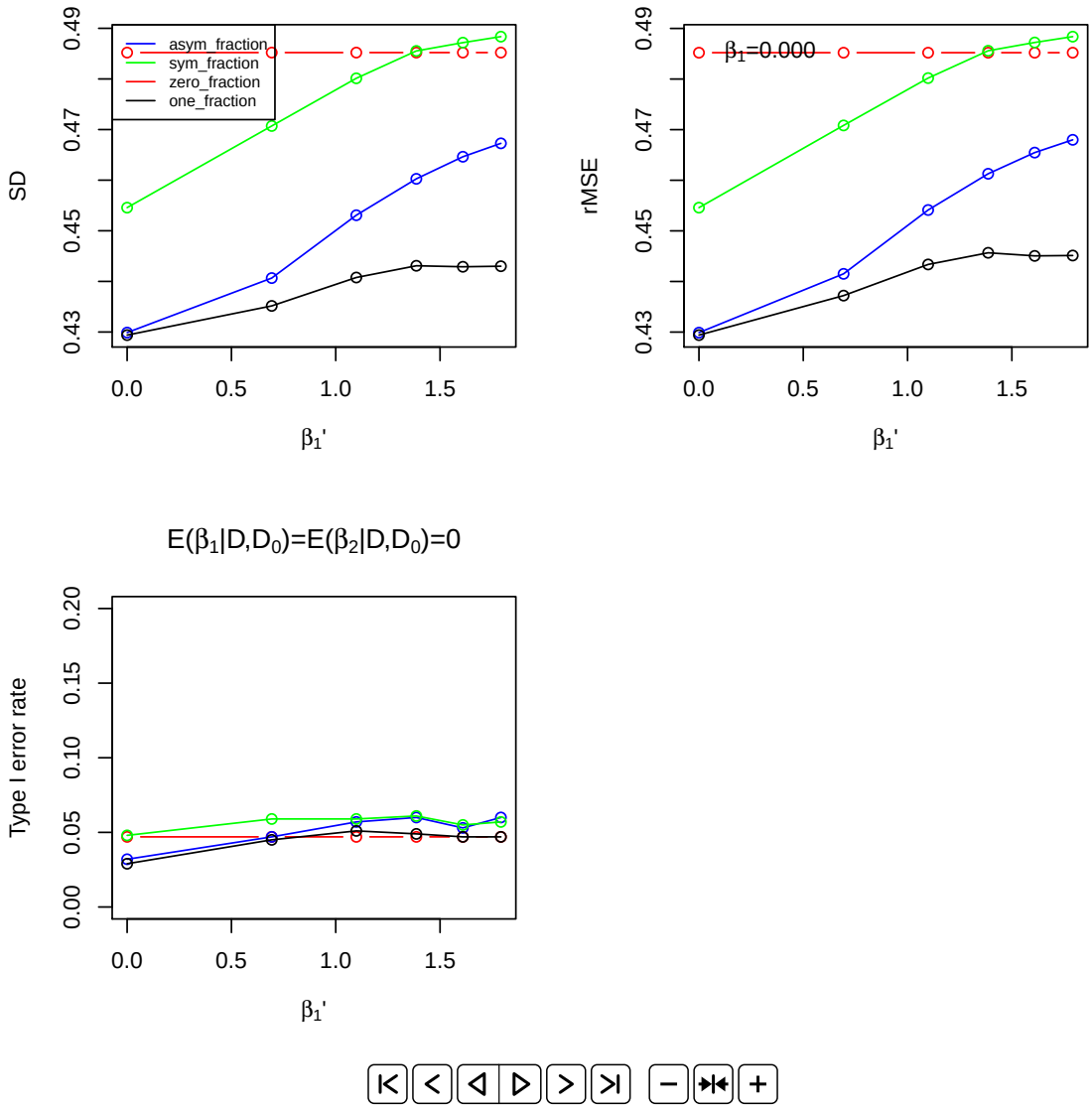

Figure 4.5: In three-arm trial with binary data, the effect of the historical treatment effect β_1' on the discounting fraction, SD and rMSE of β_2, type I error rate and power where the total number of cluster $I = 30$,the cluster size $J = 20$, and the ICC $\rho = 0.233$ in current trial and the arm size $N = 200$ in historical trial. Type I error rate and power correspond to $\beta_1 = \log(1) = 0$ and $\beta_1 = \log(4) \approx 1.386$.

Chapter 5

Computational Tools

In this chapter, we briefly describe our effort in the development of computational tools for the proposed methods. The end product is an R package, containing all necessary functions for the implementation of the described methods. At the time of completion of this dissertation, the main functions of the package have been developed and tested, but have not been released for public use.

5.1 The R Platform

Among the most widely used statistical computing platforms, R has two distinct advantages: (1) free availability of statistical computing, and (2) accommodation of user-created packages. This has led to a highly engaged and active development community. In R, packages are the basic units of collections of reusable R functions, supplemented by documentation and real data illustrations. In this dissertation, all of the computational methods discussed in the first four chapters are programmed into modularized R functions and organized into a comprehensive R package. The package allows analysts and trialists to perform Bayesian power prior analysis, as well as estimate the sample size and power for design of new studies.

A minimal R package contains three components, an "R/" directory, a DESCRIPTION file, and a NAMESPACE file (Wickham, 2015). Currently, the components may be created in a menu-driven way using Rstudio. The "R/" directory

is used to store individual R scripts, the DESCRIPTION file describes the uses and functionality of the package, and the NAMESPACE file defines which functions it avails to other packages and which functions it requires from other packages. There are usually additional tasks for completing the development of a package. For example, a "data/" directory could be created to store example data, for illustration purposes. R objects/functions, including example data, are typically documented through *roxygen2*, a LaTeX utility for R package documentation. In addition, vignette files may be created by using *Rmarkdown* and *knitr*. Vignettes are sometimes referred to a long form documentation, and provide a more detailed overview of the methods and examples of analyses.

5.2 Package overview

We developed an R package, tentatively named *pprincrt*, to cover the methods described in the previous chapters. Because the package is still being tested and has not been made public, the package and function names are subject to further changes.

The package has two main functions, *pprmodelBUGS()* and *SimPower()*. The former does the Bayesian power prior analysis for cluster randomized trials, and the latter does power calculation through simulation. The functions are able to incorporate historical trial information in both data analysis and trial design. Additionally, the package provides two utility functions: *print.pprMod()*, which prints model fitting results in an easy to read format, and *AniPlot()* for animated presentation of estimated power and type 1 error rates.

In the following subsections, we summarize the main features of the functions.

5.2.1 Function *pprmodelBUGS()*

The main utility of function *pprmodelBUGS()* is to perform the Bayesian power prior analysis. it executes a series of tasks "under the hood", outlined in order of implementation below,

- Write the file *cmodelfile.txt* into the file directory. It is a BUGS script of the GLMM model for the current trial data.
- Use function *bugs()* in the **R2OpenBUGS** package to call OpenBUGS from R and to analyze the current trial data with the model in *cmodelfile.txt* through batch mode (Spiegelhalter et al., 2007).
- Write the file *hmodelfile.txt* into the file directory. It is a BUGS script of the GLM model for the historical trial data.
- Use function *bugs()* in the **R2OpenBUGS** package to call OpenBUGS from R and to analyze the historical trial data with the model in *hmodelfile.txt* through batch mode.
- Write file *pprmodelfile.txt* into the file directory. It is a BUGS script for power prior analysis model for both current and historical trial data.

In function *pprmodelBUGS()*, the file directory is set to be *tempdir()* by default, but the users may change this through the argument *file.dir*. All files are removed after OpenBUGS is done when the argument *file.rm* is set to *TRUE*. However, the default value is set to *FALSE*. For the current trial, the data are input through the argument *cData*, while the analytical model is specified by the argument *cForm*. *cData* is a data frame, and *cForm* is a formula object. The variable names in *cData* must be the same as those in *cForm*. Besides, *hData* and *hForm* are similar

arguments to *cData* and *cForm*, but for data input and model specification of the historical trial.

The current package runs on the Window platform. The OpenBUGS directory could be specified by users through the argument *OpenBUGS.dir*, which is the directory where the OpenBUGS software is installed. By default, it is set to be *NULL*, which means that the most recent OpenBUGS version registered in the Windows registry will be used. The pseudo random number generator in OpenBUGS has 14 different internal states. Each state is 10^{12} draws apart to avoid overlap in the pseudo random number sequences (Spiegelhalter et al., 2007). The state can be pre-specified through the argument *OpenBUGS.seed* by setting its value to be an integer between 1 and 14. By default, it is set to be 1.

In function *pprmodelBUGS()*, one MCMC chain of length $2,000$ is generated. We discard the first half of the generated values and we do not use thinning. The user may change the default values for the number of the chain, the number of iteration, the number of burn-in, and the thinning parameter through arguments *nchain*, *niter*,*nburnin*,*nthin*, respectively. Two different discounting methods are allowed in the package: by the symmetric and asymmetric Kullback-Leibler divergence based discounting parameters, or by an expert proposed proportion (between 0 and 1). The discounting method is specified by the argument *weight*.

Function *pprmodelBUGS()* returns an S3 class object *pprMod* containing the posterior samples of the treatment effect, which can be further processed with the **coda** package. In addition, the object contains summaries of the posterior samples, including posterior median, standard deviation, and the highest posterior density (HPD) interval. The credible level of the HPD interval is specified by argument

cover.level. An estimate of the discounting parameter is also included in the object. Convergence diagnostic statistics are also presented, including trace plot, and the Gelman-Rubin statistic and related plots if two or more MCMC chains are run. All diagnostic plots are contained in file *graphics.pdf* under the directory specified by the argument *file.dir*.

5.2.2 Function *SimPower()*

Function *SimPower()* is used to determine the sample size in designing a new cluster randomized trial with the proposed method given a historical data or a specific historical trial parameter setting. It may also be used to assess the power of the proposed method under a pre-specified parameter setting for both current and historical trials. The desired result is specified by argument *to.do.option*. It takes value *real_SSD*, *sim_SSD* or *sim_power* for the three aforementioned aims, respectively. When argument *to.do.option* takes value *real_SSD*, function *SimPower()* executes a series of tasks "under the hood", outlined in order of implementation below,

- Identify the pre-specified historical trial data.
- Generate multiple curren trial data replicates under an experimental sample size.
- Analyze each current trial data with power prior method by borrowing information from that historical trial data, and output the HPD interval of the treatment effect.
- Use the HPD interval to make a decision whether to reject the null hypothesis of non-significant treatment effect.
- Report the proportion of rejection among the decisions as power.

If the power returned by function *SimPower()* is not the target one, then we should try another sample size and repeat the process above with function *SimPower()* until the target power is obtained, and then we could report the corresponding sample size to people. When argument *to.do.option* takes value *sim_SSD*, function *SimPower()* works similarly. The only difference is, in the first step, we should generate one historical trial data with the pre-specified historical trial parameter setting, instead. When argument *to.do.option* takes value *sim_power*, the first three steps are changed to the following two steps,

- Generate multiple pairs of current and historical data replicates under the pre-specified current and historical trial setting.
- For each pair of data replicates, analyze the current trial data with power prior method by borrowing information from the historical trial data, and output the HPD interval of the treatment effect.

In this case, the power returned by function *SimPower()* is reported to people, directly.

As shown above, multiple replicates of current trial data or historical trial data might be generated. The number of data replicates is specified by argument *Rep*. The arguments *cRdmSeed.init* and *hRdmSeed.init* are used to specify the random seeds for the generation of the first current trial data and the first historical trial data, respectively. Both random seeds increase by 1 for each of the following trial data. In concept, the computations on different data replicates are similar and independent, so they could be done on different cores of the same computer simultaneously, thereby reducing the computational time. Therefore, we allow parallel computing via the **foreach** package in this function.

Multiple files will be yielded in the file directory as a result of the computation on each data replicate, such as *'hmodelfile.txt'*, *'cmodelfile.txt'*, *'pprmodelfile.txt'* and *'graphics.pdf'*. For the concern of space, all these files are removed by default after the computation is done. This default value could be changed by setting the argument *file.rm=FALSE*. Similar to function *pprmodelBUGS()*, the same arguments are provided in function *SimPower()* to specify the OpenBUGS directory, the MCMC chain features, the discounting parameter and the HPD interval. Additionally, when *to.do.option* takes value $real\_SSD$, the historical trial data should be specified in the format of data frame through the argument *hData*, and the variable names in the data frame should be the same as those specified in the arguments *hTrtVar*, *hOutcomeVar* and *hSumVar*. When *to.do.option* takes value sim_SSD or sim_power, the historical trial parameter setting should be specified in the format of list through the argument *hSet*, and the variable names in the list should be the same as those specified in the arguments *hNarmVar*, *hNpat.armVar*, *hWthVar*, *hTrtVecVar*. Across the three aims, the current trial parameter setting should be specified in the format of list through the argument *cSet*, and the variable names in the list should be the same as those specified in the arguments *cNarmVar*, *cNcluster.armVar*, *cNpat.clusterVar*, *cBtwVar*, *cWthVar*, *cTrtVecVar*.

To facilitate the output of computational results, we developed two utility functions *print.pprMod()* and *AniPlot()*. The first function, *print.pprMod()* prints the summary statistics of posterior samples, and the second function produces animation plots, shown in the previous chapters.

5.2.3 Example 1: HPV vaccine reminder

We consider the HPV vaccine reminder trial sponsored by Merck pharmaceutical as the current trial. The outcome of interest is vaccine uptake. The trial has been described in detail in previous chapters. The structure of the data is shown in Table 5.1. On the other hand, We used the trial by Szilagyi et al. (2011) as the source of historical data. As described before, a total of $2,139$ girls, aged $11-15$ years, were recruited and randomized to two different arms. The data are summarized in Table 5.2. These two data were used to illustrate how to use the functions *pprmodelBUGS()*, *SimPower()* and *print.pprMod()*. The users can refer to the vignette file of this package for details.

5.2.4 Example 2: Animation data

The animation data is a non-real data for the illustration purpose of the function *AniPlot()*. It has four different variables, including the group variable, frame variable and X-axis and Y-axis variables. The data has 72 observations, and the group and frame variables have 2 and 6 unique values, respectively. The users can get more detailed information about the data and how to apply the function *AniPlot()* to it to create animation plot by reading the vignette file of this package.

Table 5.1: Merck HPV vaccine reminder trial

Intervention	Physician ID	Number of subjects	Number of subjects taking vaccine
0	5	14	10
0	13	22	5
0	15	50	22
0	17	11	4
0	19	20	16
0	21	15	8
0	24	10	4
0	25	23	16
0	26	3	1
0	27	55	14
0	28	2	1
1	1	7	7
1	2	24	19
1	3	16	12
1	4	9	2
1	6	1	1
1	7	12	7
1	8	11	9
1	9	16	12
1	10	45	11
1	11	37	33
1	12	12	8
1	14	2	1
1	16	22	10
1	18	2	2
1	20	7	3
1	22	19	13
1	23	8	6

Table 5.2: Szilagyi HPV vaccine reminder trial

Intervention	Number of subjects	Number of subjects taking vaccine
0	1055	453
1	1084	634

BIBLIOGRAPHY

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Beygelzimer, A., S. Kakadet, J. Langford, S. Arya, D. Mount, and S. Li (2013). Fnn: fast nearest neighbor search algorithms and applications. r package version 1.1.

Brian, N. (2010). Bayesian analysis using power priors with application to pediatric quality of care. *Journal of Biometrics & Biostatistics*.

Browne, W. J., D. Draper, et al. (2006). A comparison of bayesian and likelihood-based methods for fitting multilevel models. *Bayesian analysis 1*(3), 473–514.

Campbell, M. (2000). Cluster randomized trials in general (family) practice research. *Statistical Methods in Medical Research 9*(2), 81–94.

Campbell, M., A. Donner, and N. Klar (2007). Developments in cluster randomized trials and statistics in medicine. *Statistics in medicine 26*(1), 2–19.

Carlin, J. B. (1992). Meta-analysis for 2× 2 tables: A bayesian approach. *Statistics in medicine 11*(2), 141–158.

Chen, M.-H., J. G. Ibrahim, et al. (2006). The relationship between the power prior and hierarchical models. *Bayesian Analysis 1*(3), 551–574.

Chen, M.-H., J. G. Ibrahim, H. Amy Xia, T. Liu, and V. Hennessey (2014). Bayesian sequential meta-analysis design in evaluating cardiovascular risk in a new antidiabetic drug development program. *Statistics in medicine 33*(9), 1600–1618.

Chen, M.-H., J. G. Ibrahim, P. Lam, A. Yu, and Y. Zhang (2011). Bayesian design of noninferiority trials for medical devices using historical data. *Biometrics 67*(3), 1163–1170.

Chen, M.-H., J. G. Ibrahim, and Q.-M. Shao (2000). Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference 84*(1), 121–137.

Chen, M.-H., J. G. Ibrahim, D. Zeng, K. Hu, and C. Jia (2014). Bayesian design of superiority clinical trials for recurrent events data with applications to bleeding and transfusion events in myelodyplastic syndrome. *Biometrics 70*(4), 1003–1013.

Chen, M.-H., A. K. Manatunga, and C. J. Williams (1998). Heritability estimates from human twin data by incorporating historical prior information. *Biometrics*, 1348–1362.

Chib, S. and B. P. Carlin (1999). On mcmc sampling in hierarchical longitudinal models. *Statistics and Computing 9*(1), 17–26.

Clark, A. B. and M. O. Bachmann (2010). Bayesian methods of analysis for cluster randomized trials with count outcome data. *Statistics in medicine 29*(2), 199–209.

Congdon, P. (2007). *Bayesian statistical modelling*, Volume 704. John Wiley & Sons.

Darlington, G. A. and A. Donner (2007). Meta-analysis of community-based cluster randomization trials with binary outcomes. *Clinical Trials 4*(5), 491–498.

De Santis, F. (2006). Power priors and their use in clinical trials. *The American Statistician 60*(2), 122–129.

De Santis, F. (2007). Using historical data for bayesian sample size determination. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 170*(1), 95–113.

Donner, A. (1992). Sample size requirements for stratified cluster randomization designs. *Statistics in medicine 11*(6), 743–750.

Donner, A. (1998). Some aspects of the design and analysis of cluster randomization trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 47*(1), 95–113.

Donner, A. and N. Klar (1994). Cluster randomization trials in epidemiology: theory and application. *Journal of Statistical Planning and Inference 42*(1-2), 37–56.

Donner, A. and N. Klar (2000). *Design and analysis of cluster randomization trials in health research*.

Donner, A. and N. Klar (2002). Issues in the meta-analysis of cluster randomized trials. *Statistics in medicine 21*(19), 2971–2980.

Donner, A., G. Piaggio, and J. Villar (2001). Statistical methods for the meta-analysis of cluster randomization trials. *Statistical Methods in Medical Research 10*(5), 325–338.

Donner, A., G. Piaggio, and J. Villar (2003). Meta-analyses of cluster randomization trials power considerations. *Evaluation & the health professions 26*(3), 340–351.

Duan, Y. (2005). *A modified bayesian power prior approach with applications in water quality evaluation*. Ph. D. thesis, Virginia Polytechnic Institute and State University.

Duan, Y., K. Ye, and E. P. Smith (2006). Evaluating water quality using power priors to incorporate historical information. *Environmetrics 17*(1), 95–106.

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing 7*(1), 57–68.

Gelman, A. et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis 1*(3), 515–534.

Goodman, S. N. and J. T. Sladky (2005). A bayesian approach to randomized controlled trials in children utilizing information from adults: the case of guillain-barre. *Clinical Trials 2*(4), 305–310.

Guyatt, G. H., D. L. Sackett, J. C. Sinclair, R. Hayward, D. J. Cook, R. J. Cook, E. Bass, H. Gerstein, B. Haynes, A. Holbrook, et al. (1995). Users' guides to the medical literature: Ix. a method for grading health care recommendations. *Jama 274*(22), 1800–1804.

Hampson, L. V., J. Whitehead, D. Eleftheriou, and P. Brogan (2014). Bayesian methods for the design and interpretation of clinical trials in very rare diseases. *Statistics in medicine 33*(24), 4186–4201.

Hemming, K., A. J. Girling, A. J. Sitch, J. Marsh, and R. J. Lilford (2011). Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. *BMC medical research methodology 11*(1), 1.

Hobbs, B. P. and B. P. Carlin (2007). Practical bayesian design and analysis for drug and device clinical trials. *Journal of biopharmaceutical statistics 18*(1), 54–80.

Hobbs, B. P., B. P. Carlin, S. J. Mandrekar, and D. J. Sargent (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics 67*(3), 1047–1056.

Hobbs, B. P., D. J. Sargent, and B. P. Carlin (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian analysis (Online) 7*(3), 639.

Hsu, J. (1996). *Multiple comparisons: theory and methods*. CRC Press.

Ibrahim, J. G. and M.-H. Chen (2000). Power prior distributions for regression models. *Statistical Science*, 46–60.

Ibrahim, J. G., M.-H. Chen, Y. Gwon, and F. Chen (2015). The power prior: theory and applications. *Statistics in medicine 34*(28), 3724–3749.

Ibrahim, J. G., M.-H. Chen, H. A. Xia, and T. Liu (2012). Bayesian meta-experimental design: Evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes. *Biometrics 68*(2), 578–586.

Kerry, S. M. and J. M. Bland (1998). Statistics notes: Sample size in cluster randomisation. *Bmj 316*(7130), 549.

Kong, S.-H., C. W. Ahn, and S.-H. Jung (2003). Sample size calculation for dichotomous outcomes in cluster randomization trials with varying cluster size. *Drug information journal 37*(1), 109–114.

Laopaiboon, M. (2003). Meta-analyses involving cluster randomization trials: a review of published literature in health care. *Statistical methods in medical research 12*(6), 515–530.

Lunn, D., D. Spiegelhalter, A. Thomas, and N. Best (2009). The bugs project: Evolution, critique and future directions. *Statistics in medicine 28*(25), 3049–3067.

Manatunga, A. K., M. G. Hudgens, and S. Chen (2001). Sample size estimation in cluster randomized studies with varying cluster size. *Biometrical Journal 43*(1), 75–86.

Matteucci, M. and B. P. Veldkamp (2015). The approach of power priors for ability estimation in irt models. *Quality & Quantity 49*(3), 917–926.

Pezeshk, H. (2003). Bayesian techniques for sample size determination in clinical trials: a short review. *Statistical Methods in Medical Research 12*(6), 489–504.

Rietbergen, C., I. Klugkist, K. J. Janssen, K. G. Moons, and H. J. Hoijtink (2011). Incorporation of historical data in the analysis of randomized therapeutic trials. *Contemporary clinical trials 32*(6), 848–855.

Robert, C. and G. Casella (2009). *Introducing Monte Carlo Methods with R*. Springer Science & Business Media.

Rotondi, M. A. and A. Donner (2009). Sample size estimation in cluster randomized educational trials: An empirical bayes approach. *Journal of Educational and Behavioral Statistics 34*(2), 229–237.

Schoenfeld, D. A., H. Zheng, and D. M. Finkelstein (2009). Bayesian design using adult data to augment pediatric trials. *Clinical Trials 6*(4), 297–304.

Shao, K. (2012). A comparison of three methods for integrating historical information for bayesian model averaged benchmark dose estimation. *Environmental toxicology and pharmacology 34*(2), 288–296.

Shuster, J. J., L. S. Jones, and D. A. Salmon (2007). Fixed vs random effects meta-analysis in rare event studies: The rosiglitazone link with myocardial infarction and cardiac death. *Statistics in medicine 26*(24), 4375–4385.

Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn (2007). Openbugs user manual, version 3.0. 2. *MRC Biostatistics Unit, Cambridge*.

Spiegelhalter, D. J. (2001). Bayesian methods for cluster randomized trials with continuous responses. *Statistics in medicine 20*(3), 435–452.

Stamey, J. D., F. Natanegara, and J. W. Seaman Jr (2013). Bayesian sample size determination for a clinical trial with correlated continuous and binary outcomes. *Journal of biopharmaceutical statistics 23*(4), 790–803.

Stryhn, H., J. Sanchez, P. Morley, C. Booker, and I. Dohoo (2006). Interpretation of variance parameters in multilevel poisson regression models. In *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics*.

Szilagyi, P. G., S. G. Humiston, S. Gallivan, C. Albertin, M. Sandler, and A. Blumkin (2011). Effectiveness of a citywide patient immunization navigator program on improving adolescent immunizations and preventive care visit rates. *Archives of pediatrics & adolescent medicine 165*(6), 547–553.

Turner, R. M., R. Z. Omar, and S. G. Thompson (2001). Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in medicine 20*(3), 453–472.

Turner, R. M., S. G. Thompson, and D. J. Spiegelhalter (2005). Prior distributions for the intracluster correlation coefficient, based on multiple previous estimates, and their application in cluster randomized trials. *Clinical Trials 2*(2), 108–118.

Turner, R. M., A. Toby Prevost, and S. G. Thompson (2004). Allowing for imprecision of the intracluster correlation coefficient in the design of cluster randomized trials. *Statistics in medicine 23*(8), 1195–1214.

Viele, K., S. Berry, B. Neuenschwander, B. Amzal, F. Chen, N. Enas, B. Hobbs, J. G. Ibrahim, N. Kinnersley, S. Lindborg, et al. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical statistics 13*(1), 41–54.

Wang, F. and A. E. Gelfand (2002). A simulation-based approach to bayesian sample size determination for performance under a given model and for separating models. *Statistical Science*, 193–208.

Wang, Q., S. R. Kulkarni, and S. Verdu (2009). Divergence estimation for multidimensional densities via-nearest-neighbor distances. *IEEE Transactions on Information Theory 55*(5), 2392–2405.

Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *Journal of the Royal Statistical Society: Series D (The Statistician) 46*(2), 185–191.

Wickham, H. (2015). *R packages: organize, test, document, and share your code.* " O'Reilly Media, Inc.".

Zhang, H. (2010). *Bayesian Power Prior Analysis and Its Application to Operational Risk and Rasch Model.* ERIC.

CURRICULUM VITAE

Shan Xiao

EDUCATION

- Ph.D. in Biostatistics, Minor in Bioinformatics, Indiana University, Indianapolis, IN, 2017
- M.S. in Applied Statistics, Purdue University, Indianapolis, IN, 2014
- M.S. in Epidemiology & Biostatistics, Southeast University, Nanjing, China, 2011
- B.S. in Applied Mathematics, Southeast University, Nanjing, China, 2008

WORKING EXPERIENCE

- Summer Intern, CDER/FDA, Silver Spring, MD, June 2016-Sep 2016
- Research Assistant, Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN, July 2014-May 2016

SELECT PUBLICATIONS

- Huang, D.L., Xiao, S., Dang, Q.Y., Tsong, Y. (2017). Concentration-QTc Modeling: A Simulation Study. *Submitted*.
- Zimet, G.D., Dixon, B.E., Xiao, S., Tu, W.Z., Kulkarni, A., Downs, S.M. (2017). Effects of Automated Physician Reminders and Recommendation Scripts on Initiation of HPV Vaccination. *Submitted*.

- Zimet, G.D., Dixon, B.E., Xiao, S., Tu, W.Z., Linday, B., Sheley, M.E., Dugan, T., Church, A., Downs, S.M. (2017). Physician Computerized Decision-Support Systems (CDSS) Prompts and Administration of 2nd and 3rd Doses of HPV Vaccine: A Randomized Clinical Trial. *Submitted*.
- Stupiansky, N., Liau, A., Rosenberger, J., Rosenthal, S., Xiao, S., *et al* (2017). Young Men's Disclosure of Same Sex Behaviors to Health Care Providers and the Impact on Health: Results from a U.S. National Sample of Young Men who have Sex with Men. *AIDS patient care and STDS. Accepted*.
- Aalsma, M., Gilbert, A., Xiao, S., Rickert V. (2016). Parent and Adolescent Barriers and Facilitators to Adolescent Preventive Health Care Utilization. *The Journal of pediatrics.***169**: 140-145
- Liu, P., Xiao, S., Shi, Z., *et al* (2011). Bayesian evaluation of the HIV antibody screening strategy of duplicate ELISA in Xuzhou blood center, China. *Transfusion.* **51(4)**: 793–798.
- Liu, P., Yang, H., Qiang, L., Jin, H., Xiao, S., Shi, Z. (2010). Evaluation of 30 commercial assays for the detection of antibodies to HIV in China using classical and Bayesian statistics. *Journal of virological methods.* **170(1)**: 73–79.

PRESENTATIONS

- A simulation study to evaluate the performance of different models in concentration-QT analysis. Oral presentation at CDER/FDA, August 2016. Silver Spring, MD.